

Het Geheugen van Vlaanderen

Pieter Moors, Rudy Dekeerschieter, & Johan Wagemans

Laboratorium voor Experimentele Psychologie (KU Leuven)

Introductie

Het Geheugen van Vlaanderen is een massa-experiment met als doel een zicht te krijgen op het kortetermijngeheugen van de Vlaming. Het omvat een reeks geheugentestjes die afgenomen werden in het kader van de Dag van de Wetenschap. In totaal waren er 9 verschillende testen beschikbaar die telkens hetzelfde stramien volgden. Bij de start van elke test worden 20 afbeeldingen gepresenteerd die de deelnemer zo goed mogelijk diende te onthouden. In een tweede fase werden er opnieuw 20 afbeeldingen gepresenteerd. De helft hiervan waren afbeeldingen uit de eerste reeks afbeeldingen. Voor elk van de 20 afbeeldingen diende de deelnemer aan te geven of de aangeboden afbeelding al dan niet in de vorige reeks gepresenteerd werd. 2 van de 9 testen waren basistesten (eentje met woorden en eentje met beelden), die aan elke deelnemer aangeboden werden. De overige 7 testen (schoenen, auto's, kinderen, categorieën, frequentie, vormpjes, gezichten) konden na deelname aan de eerste 2 testjes geselecteerd worden. In dit document gaan we dieper in op de resultaten die deze geheugentesten hebben opgeleverd. We introduceren eerst wat algemene cijfers van deze data set.

Structuur van de dataset

De dataset bestaat momenteel uit 18174 unieke deelnemers, waarvan 38% mannen. -18174 personen startten de test, maar werkten hem vervolgens niet af. Volgende tabellen en figuren geven een overzicht van de data in functie van geslacht, leeftijd, en provincie.

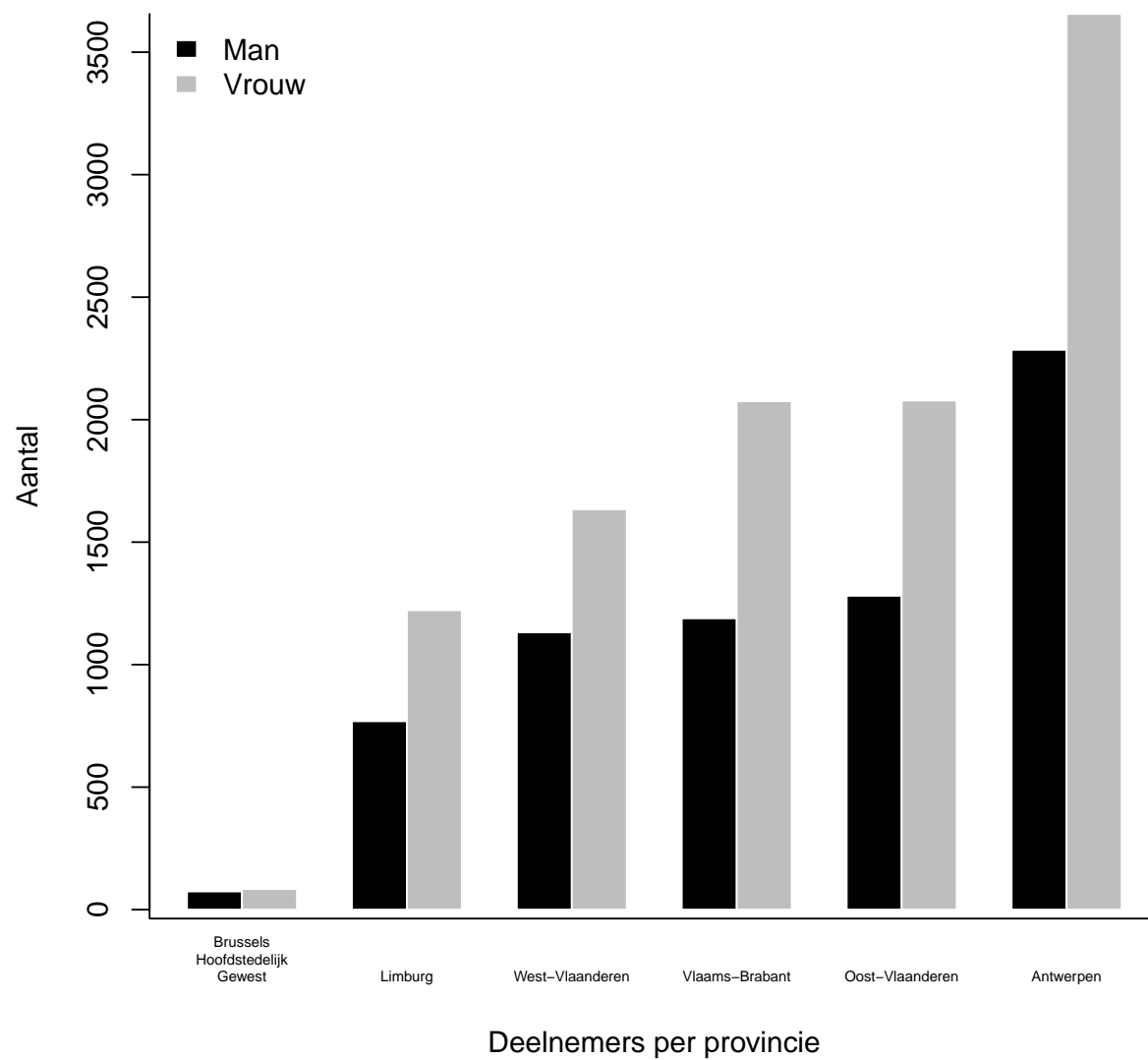
Zoals reeds gezegd is ongeveer tweederde van de deelnemers vrouw.

Geslacht	Aantal	Percentage
Man	6839	38
Vrouw	11335	62

De deelnamepercentages in functie van provincie zijn niet volledig gelijk. Het Brussels Hoofdstedelijk Gewest is sterk ondergerepresenteerd. Antwerpen heeft een iets hoger deelnamepercentage en het Limburgs deelnamepercentage ligt dan weer wat lager.

Provincie	Aantal	Percentage
Antwerpen	6345	35
Brussels Hoofdstedelijk gewest	162	1
Limburg	2045	11
Oost-Vlaanderen	3442	19
Vlaams-Brabant	3331	18
West-Vlaanderen	2849	16

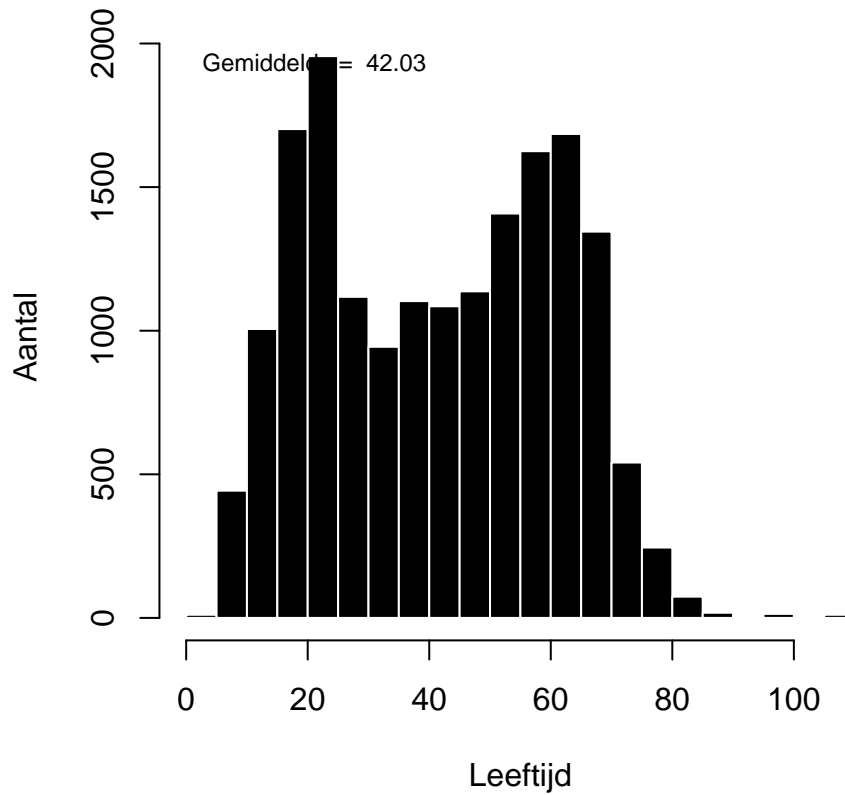
Het staafdiagram hieronder geeft de verdeling weer in functie van leeftijd en geslacht. We zien dat in quasi elke provincie er meer vrouwen deelnemen dan mannen.



In termen van leeftijd, zien we dat de grootste groep van deelnemers (ongeveer de helft) tussen de 46 en 69 jaar oud is. De steekproef bestaat momenteel dus vooral uit iets oudere vrouwen. 695 deelnemers geen leeftijd opgegeven hebben. Opvallend is dat 84% van deze deelnemers vrouw is. In onderstaande tabel zijn alle deelnemers met leeftijd 0 verwijderd voor de percentages berekend werden.

Leeftijd	Aantal	Percentage
<= 12	857	5
13-17	992	5
18-22	2168	12
23-34	2983	16
35-45	2368	13
46-57	3156	17
58-69	3823	21
>= 70	1132	6

Volgend histogram visualiseert de leeftijdsverdeling van de deelnemers.



Tot slot bekijken we nog even de verdeling over de verschillende testjes. Zoals verwacht hebben de meeste mensen deelgenomen aan de basistesten. 5 à 13% van de deelnemers legt ook één of meerdere optionele testjes af. De meeste deelnemers leggen 1, 2 of 3 testjes af.

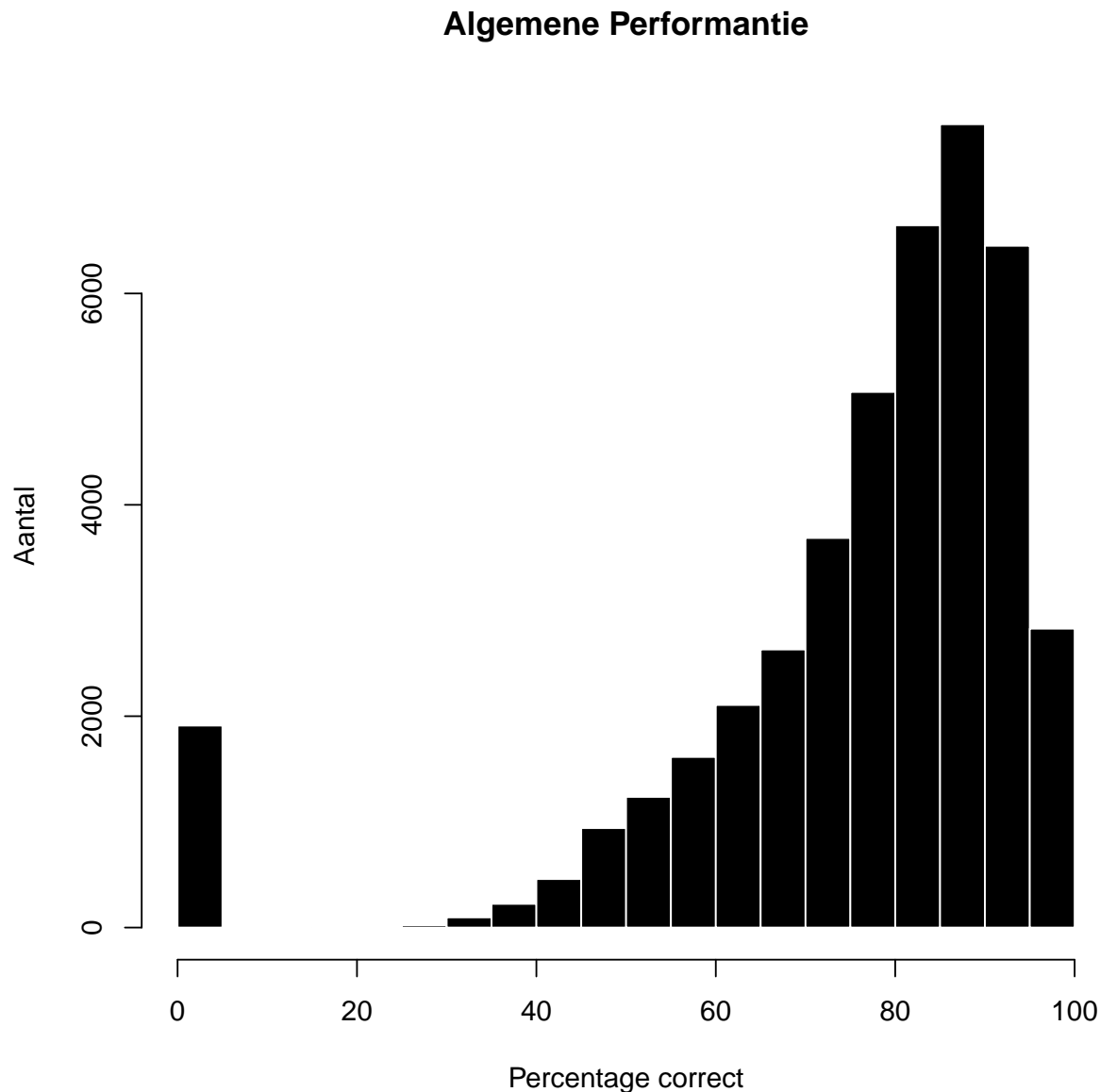
Test	Aantal
beelden (scenes)	17102
woorden (basis)	16994
beelden (gezichten)	2286
beelden (autos)	1681
beelden (schoenen)	1315
beelden (kinderen)	1222
beelden (vormpjes)	1141
woorden (frequentie)	915
woorden (categorie)	875

Aantal tests afgelegd	Aantal
1	2229
2	12090
3	1791
4	862
5	390
6	207
7	128
8	80
9	397

Een eerste analyse

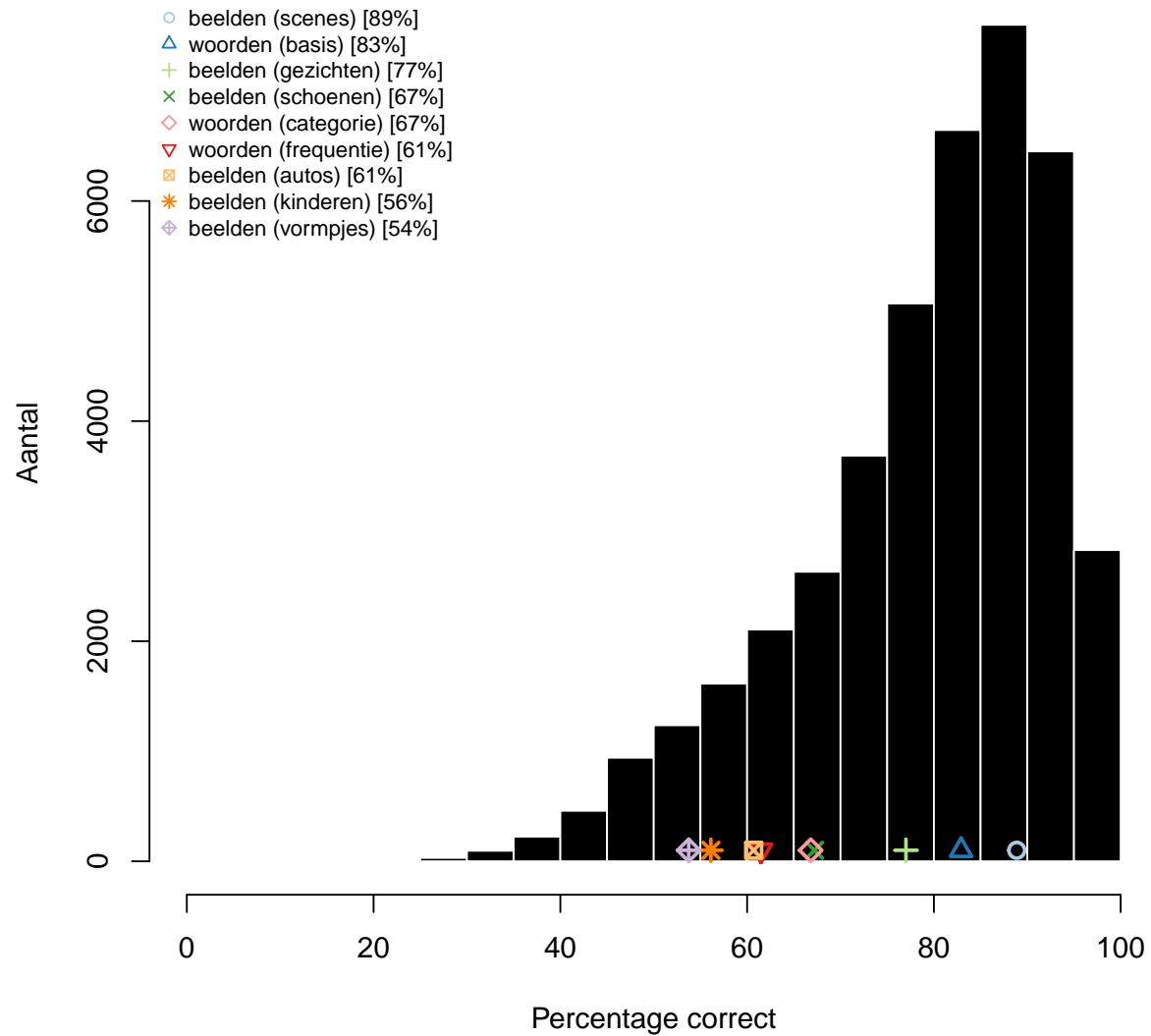
Zonder na te gaan of er verschillen zijn in de performantie tussen verschillende geheugentesten, bekijken we de algemene performantie en de performantie in functie van geslacht, leeftijd, en provincie.

Volgend histogram geeft de verdeling van de scores van alle deelnemers voor alle testen weer. Wat dadelijk opvalt is dat er een distincte groep van deelnemers is die sterk onder het gemiddelde scoort. We vermoeden dat deze deelnemers (1912 personen) de taak omgekeerd hebben en waar ze ja dienden te antwoorden consistent nee geantwoord hebben en omgekeerd. Voor verdere analyses sluiten we deze groep van deelnemers uiteraard uit.

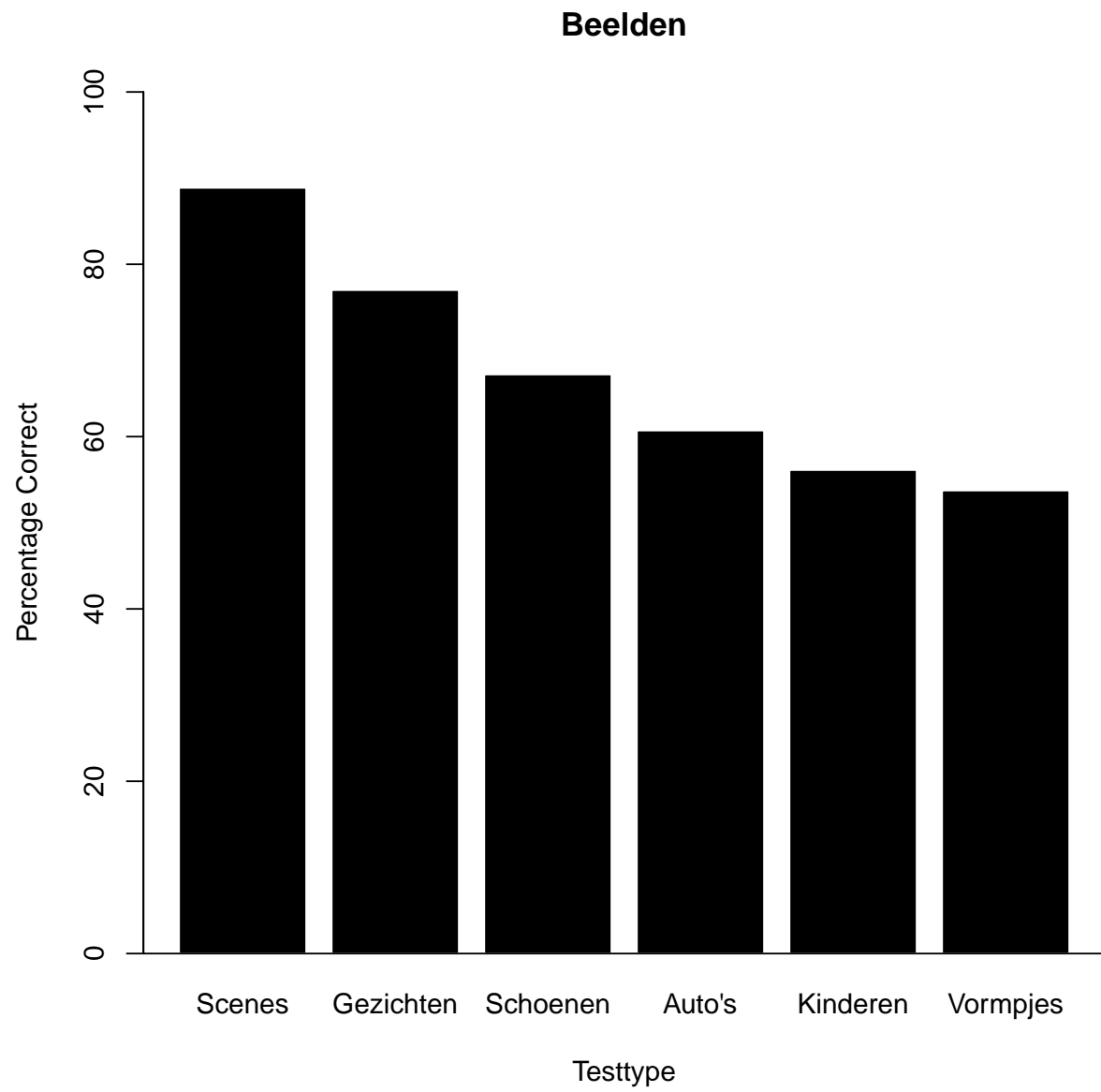


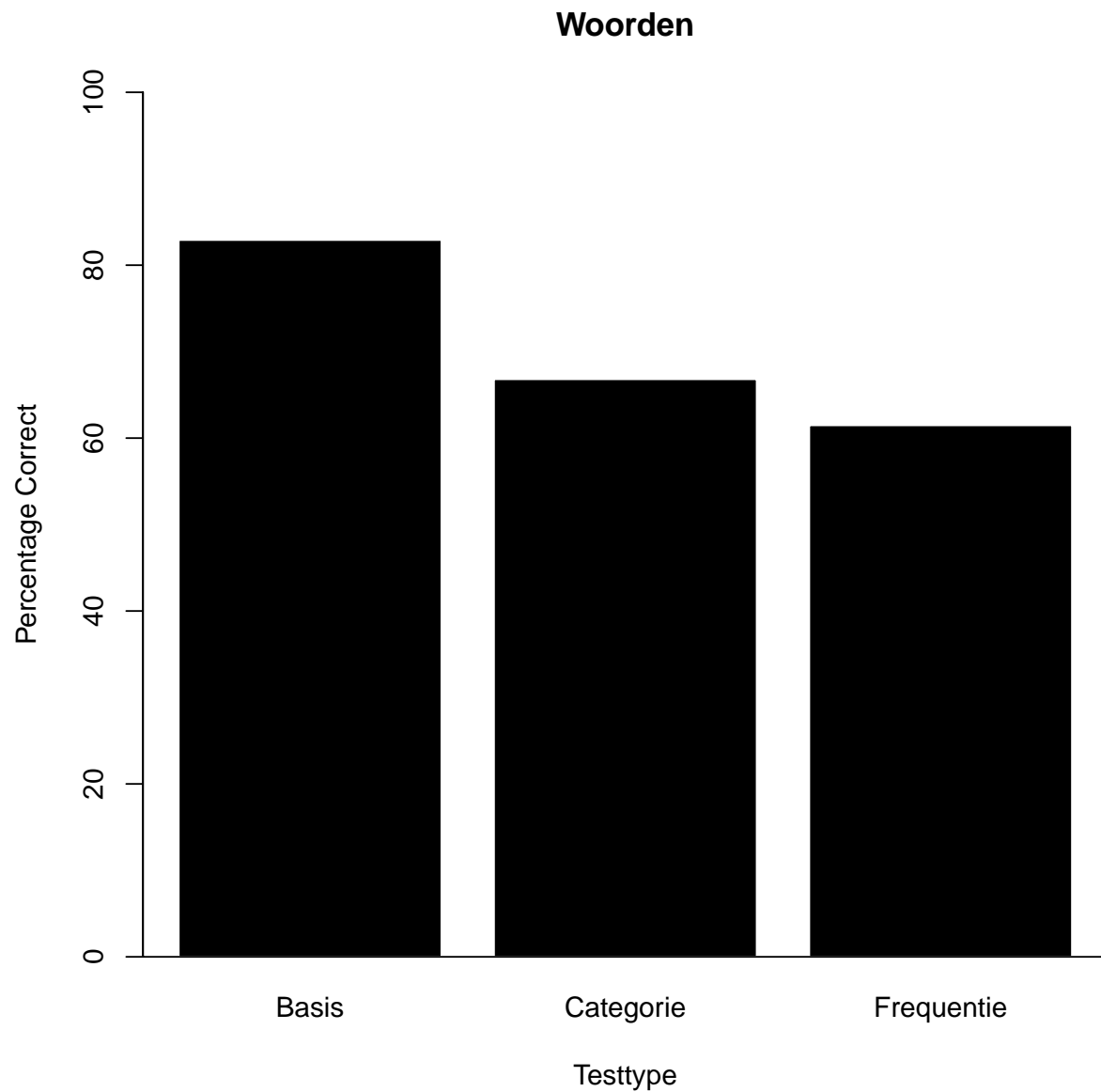
Hetzelfde histogram wordt hieronder weergegeven (uitgezuiverd voor de zeer lage performanties). De symbolen op het histogram geven de gemiddelde performantie op elke test afzonderlijk weer. Beide basistesten hebben de hoogste gemiddelde performantie. Zoals verwacht is de test met de abstracte vormen de moeilijkste test en hier scoren deelnemers bijna niet beter dan wanneer ze volledig willekeurig zouden antwoorden.

Algemene Performantie

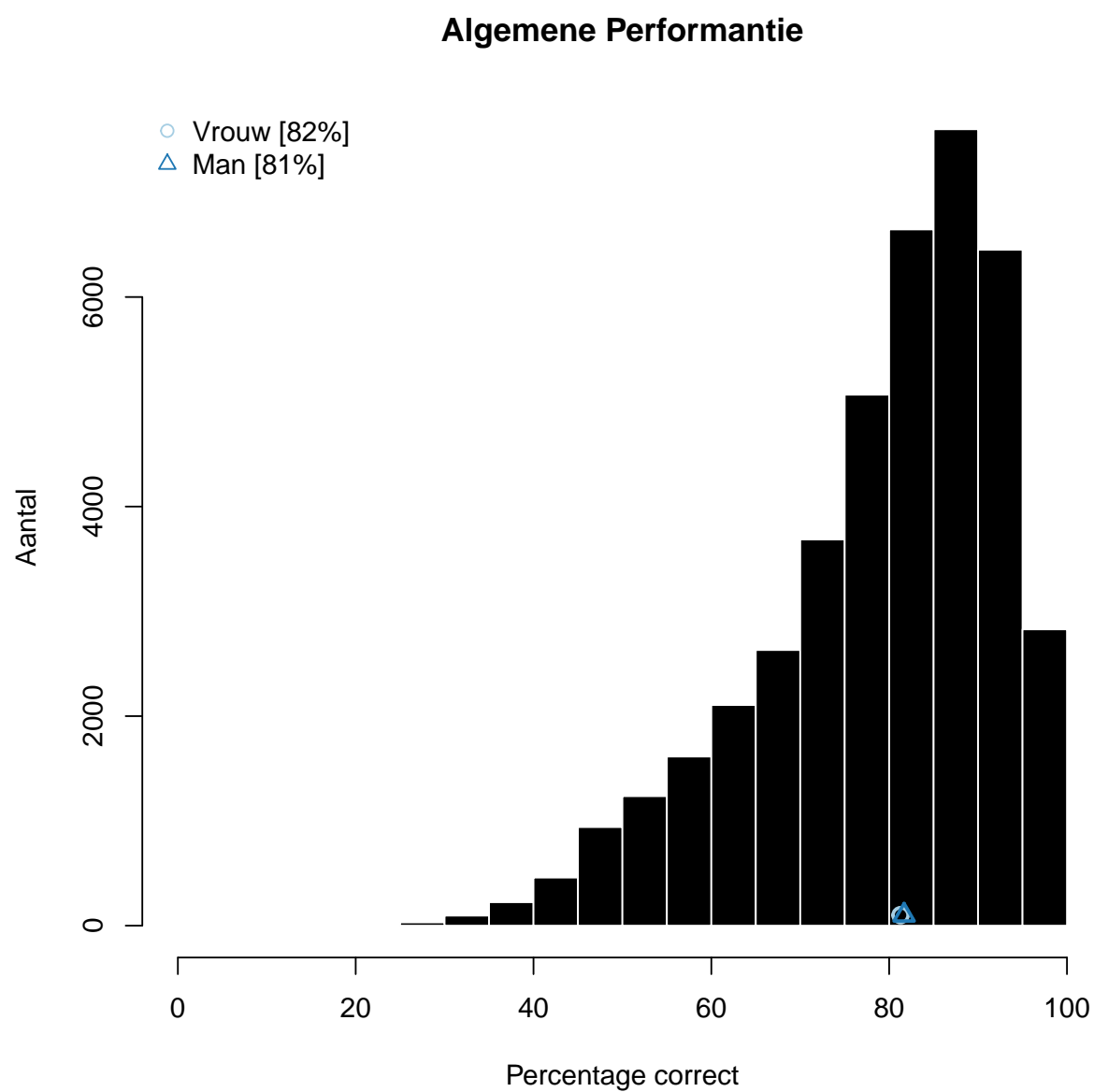


Onderstaande staafdiagrammen geven de gemiddelde percentages correct weer voor de verschillende testen met beelden en die met woorden.

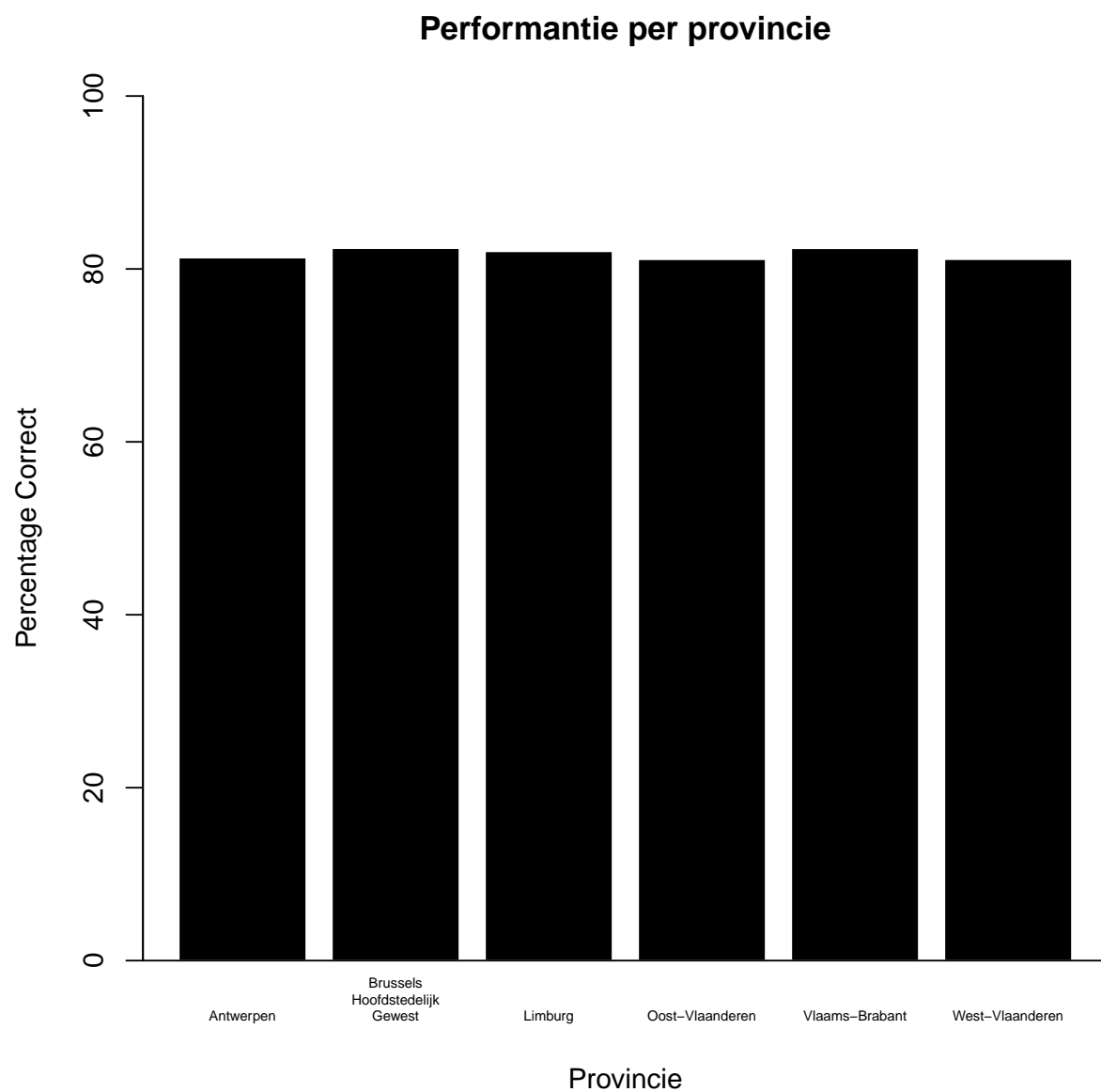




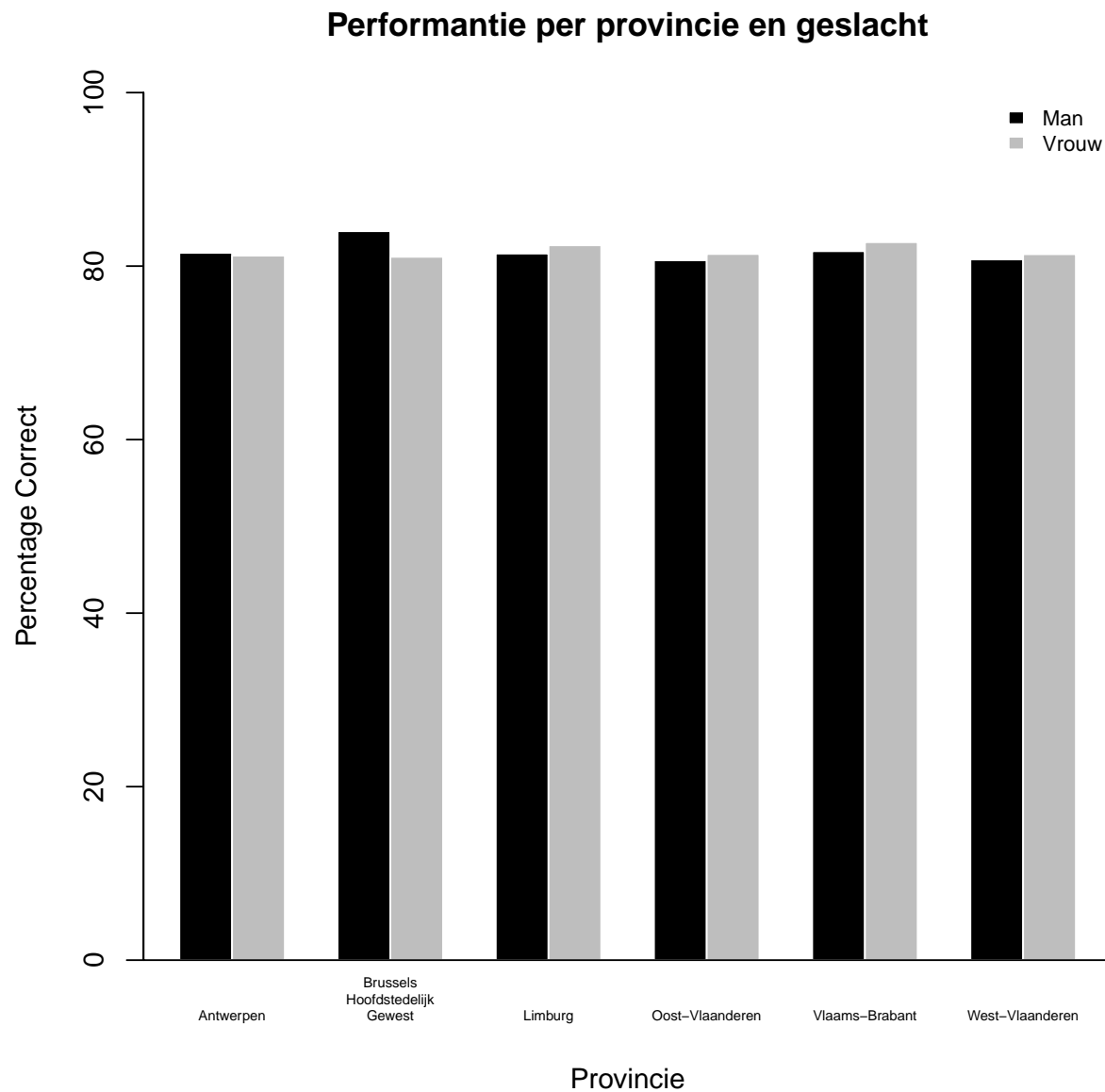
Hieronder geven we opnieuw het histogram van performanties weer, maar nu met de verschillen in functie van geslacht. Over alle testen heen scoren vrouwen gemiddeld 1% beter dan mannen. Hoewel deze trend zich lijkt voort te zetten in de gehele dataset, gaan de werkelijke verschillen steeds over ongeveer 1 à 2%.



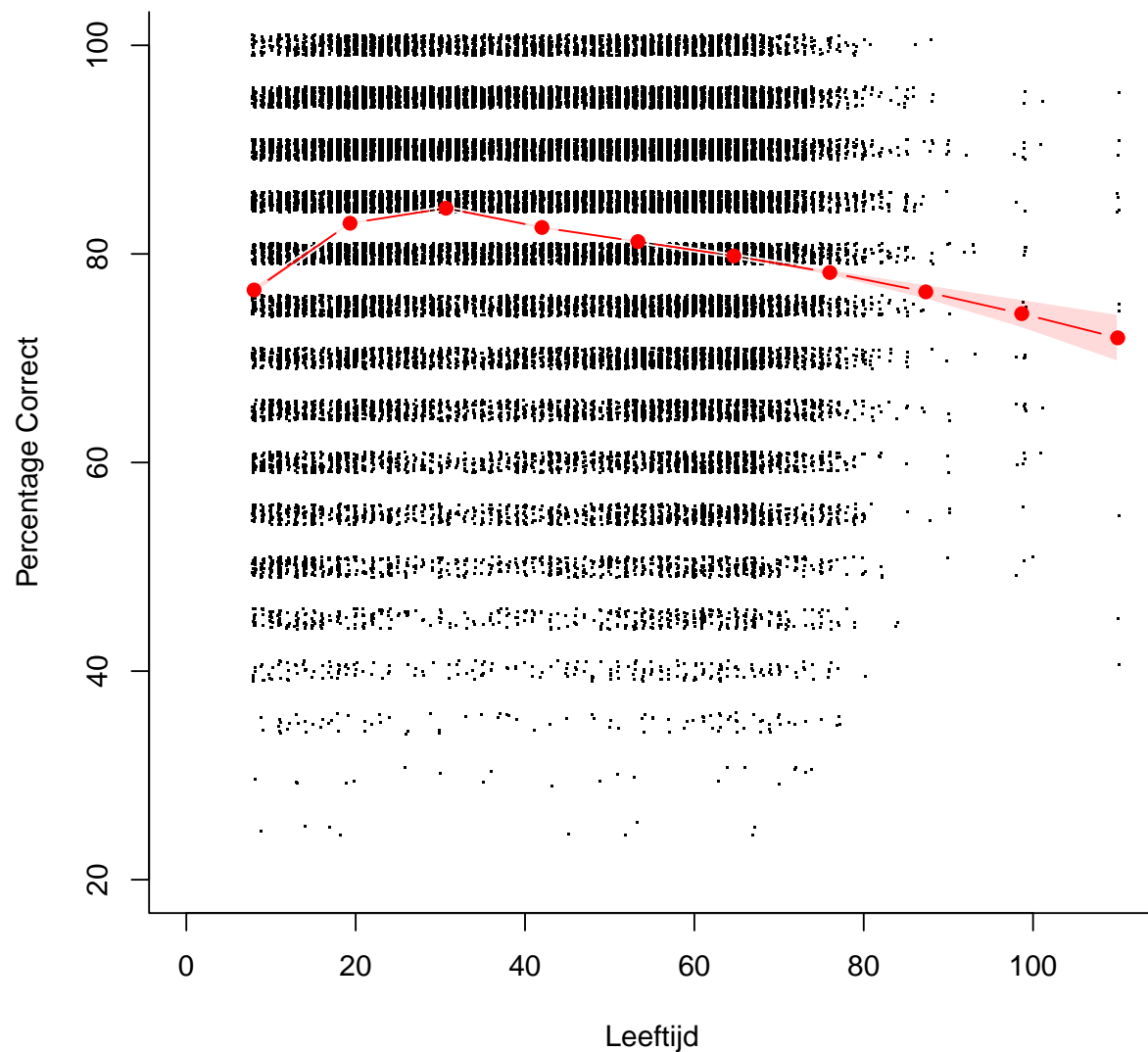
Indien we de data opsplitsen per provincie, zien we dat er geen indicatie is voor een verschil tussen provincies.



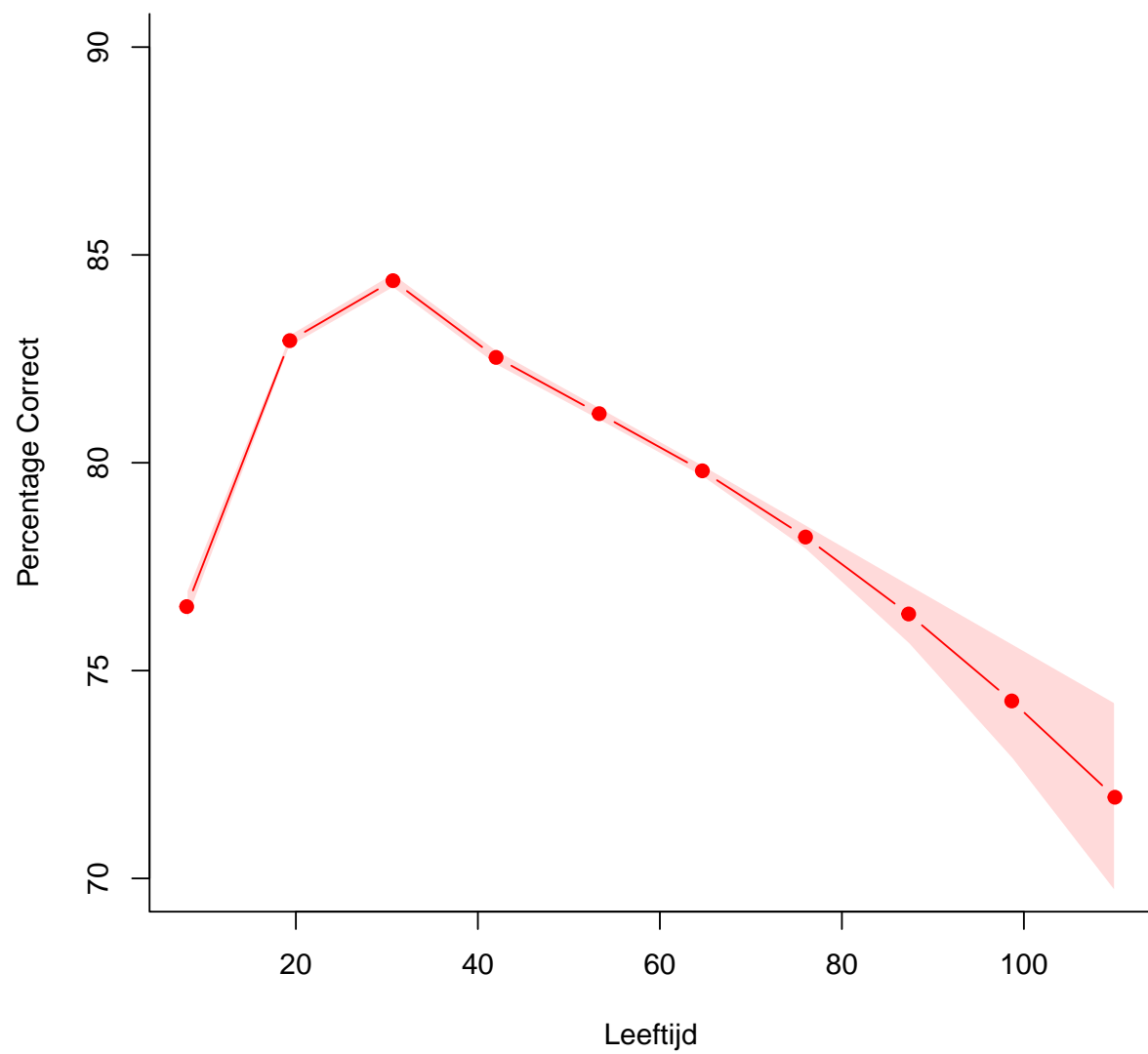
We kunnen de data ook verder opsplitsen naar provincie en geslacht, maar het wordt opnieuw duidelijk dat er, als er al verschillen zijn, deze minimaal zijn.



Een interessanter patroon tekent zich af wanneer we de performantie afzetten tegen de leeftijd van de deelnemers. Hoewel uit de ruwe data (kleine puntjes) op het eerste zicht niet zo veel af te leiden valt, is het duidelijk dat de rode lijn (op te vatten als een lijn die het verband tussen beide variabelen zo goed mogelijk beschrijft) een curvilineair verband weergeeft tussen leeftijd en performantie. Jonge deelnemers beginnen reeds op een relatief hoog percentage correct waarna dit verder stijgt tot ongeveer 30 jaar en hierna gestaag begint te dalen tot het terug afvlakt rond de 80 jaar.

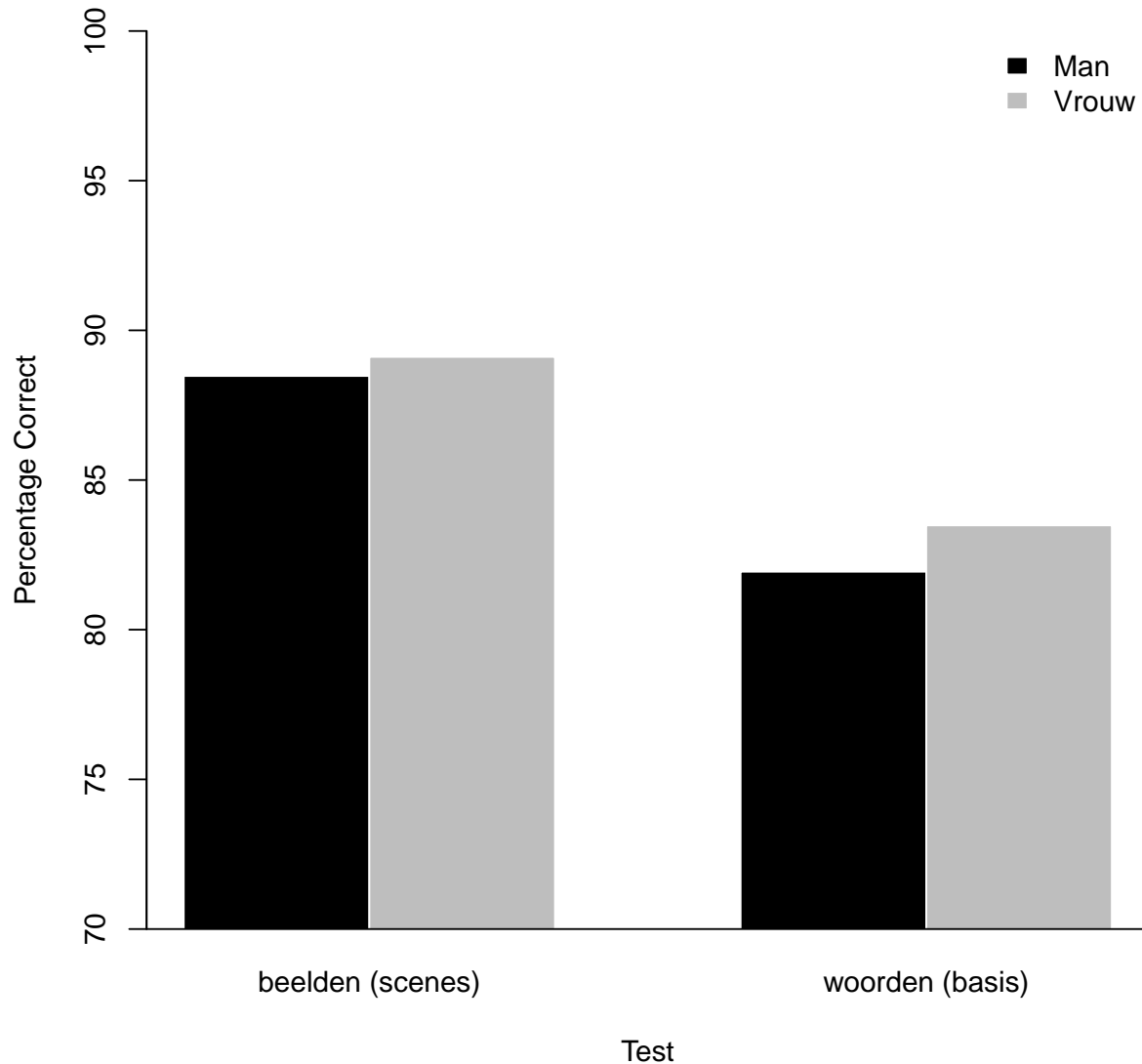


Volgende curve zoomt specifiek in op de regio waarin het bovenstaand curvilineair verband afgebeeld is. De transparante rode regio geeft de onzekerheid weer van de schattingen van de rode datapunten. Omdat we relatief weinig data hebben voor deelnemers ouder dan 80 jaar is het logisch dat de onzekerheid hier groter wordt.

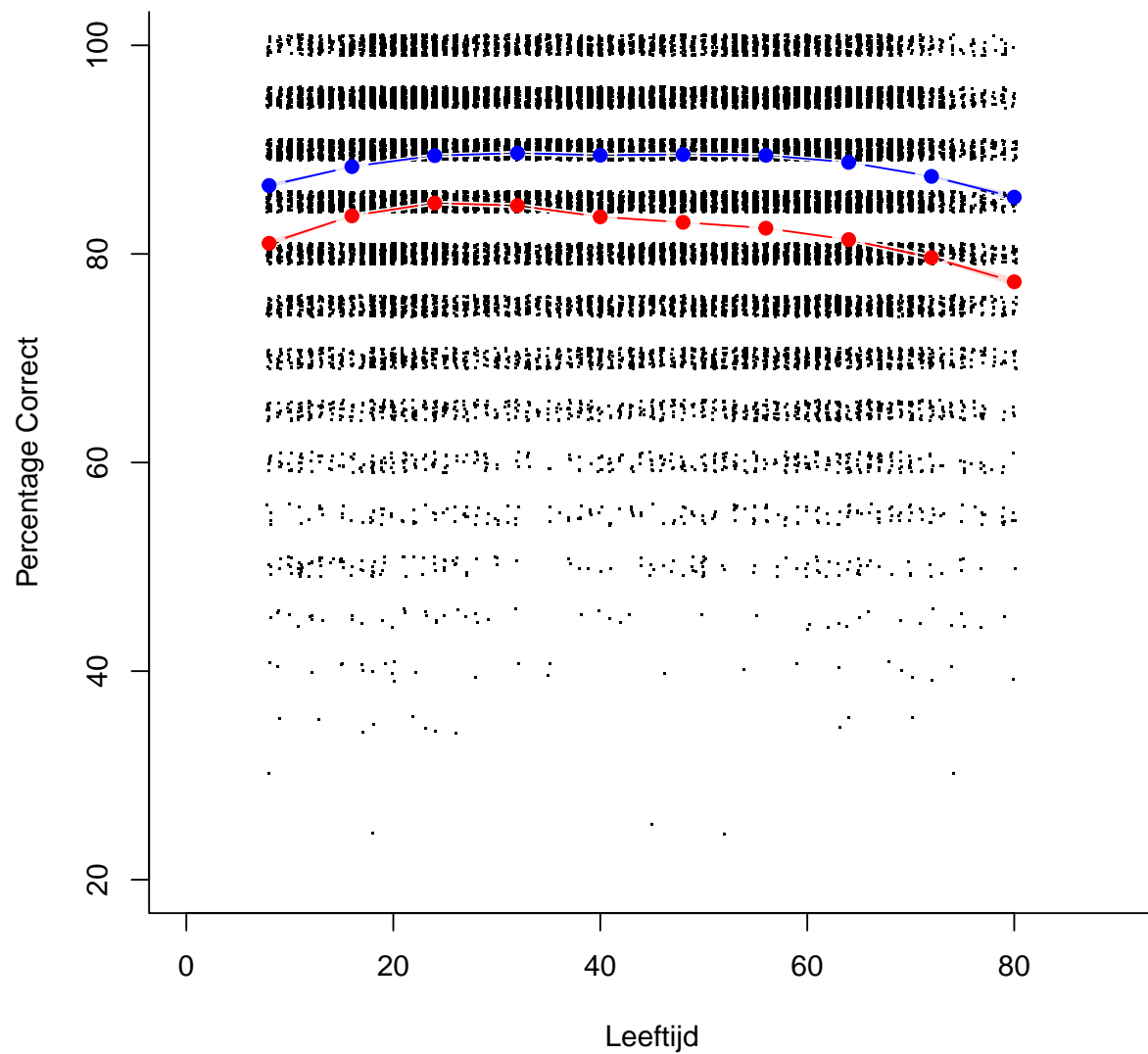


Analyse van de twee hoofdtesten

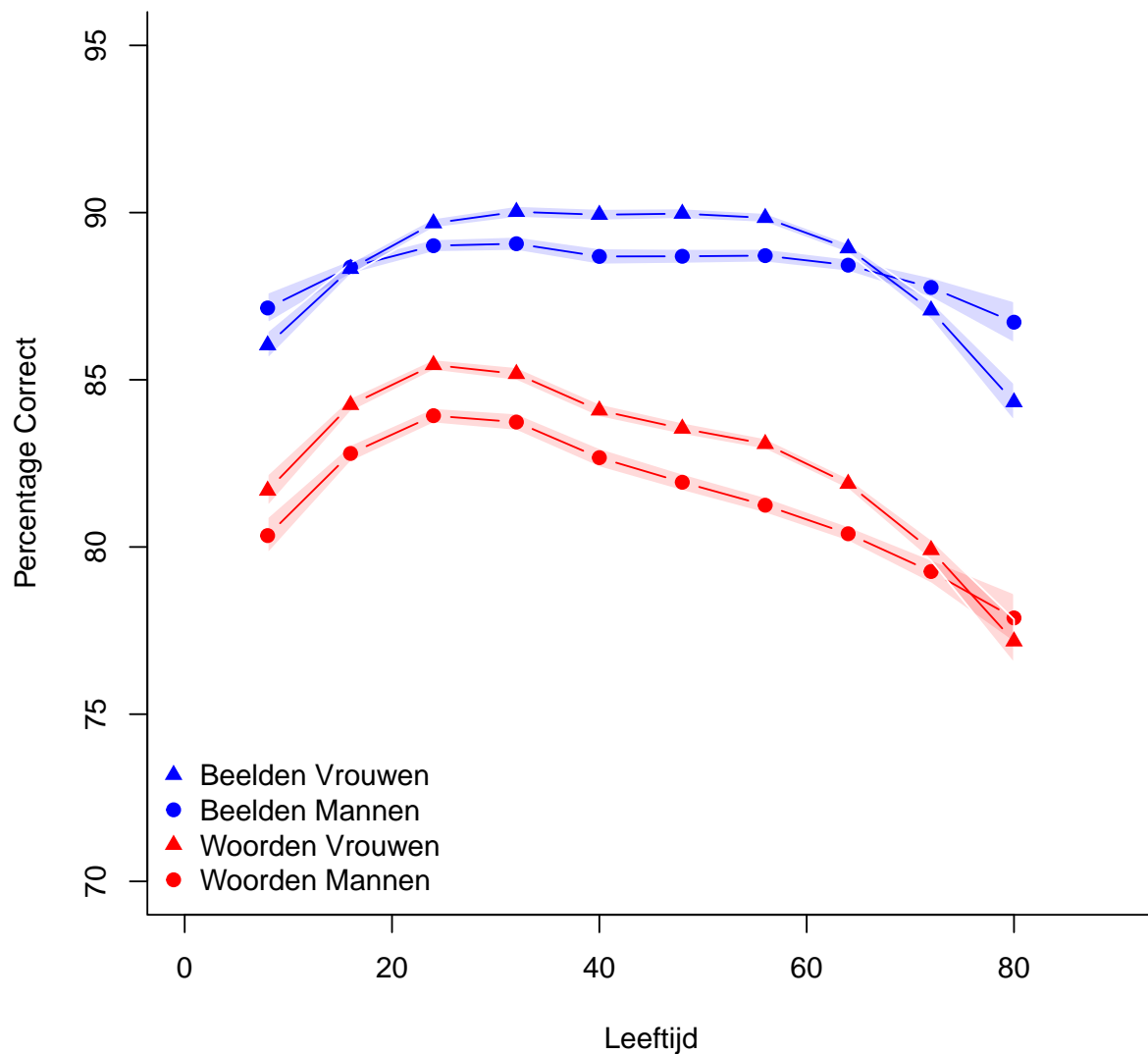
Vervolgens gaan we wat dieper in op de performantie op de twee basistesten, die met beelden en die met woorden. Zoals eerder gezegd, is de performantie gemiddeld genomen beter op de beeldentest dan op de woordentest. Onderstaand staafdiagram geeft dit weer. Opnieuw zijn vrouwen lichtjes beter dan mannen (het absoluut verschil is echter ongeveer 1%).



Indien we opnieuw in functie van leeftijd naar de data kijken, zien we eenzelfde curvilineair patroon dat zich aftekent (Noteer dat we de deelnemers ouder dan 80 nu ook niet meer beschouwen). Op alle leeftijdsniveaus is de performantie beter voor beelden dan voor woorden. Het lijkt wel het geval te zijn dat het verval minder sterk is voor beelden dan voor woorden bij hogere leeftijd.

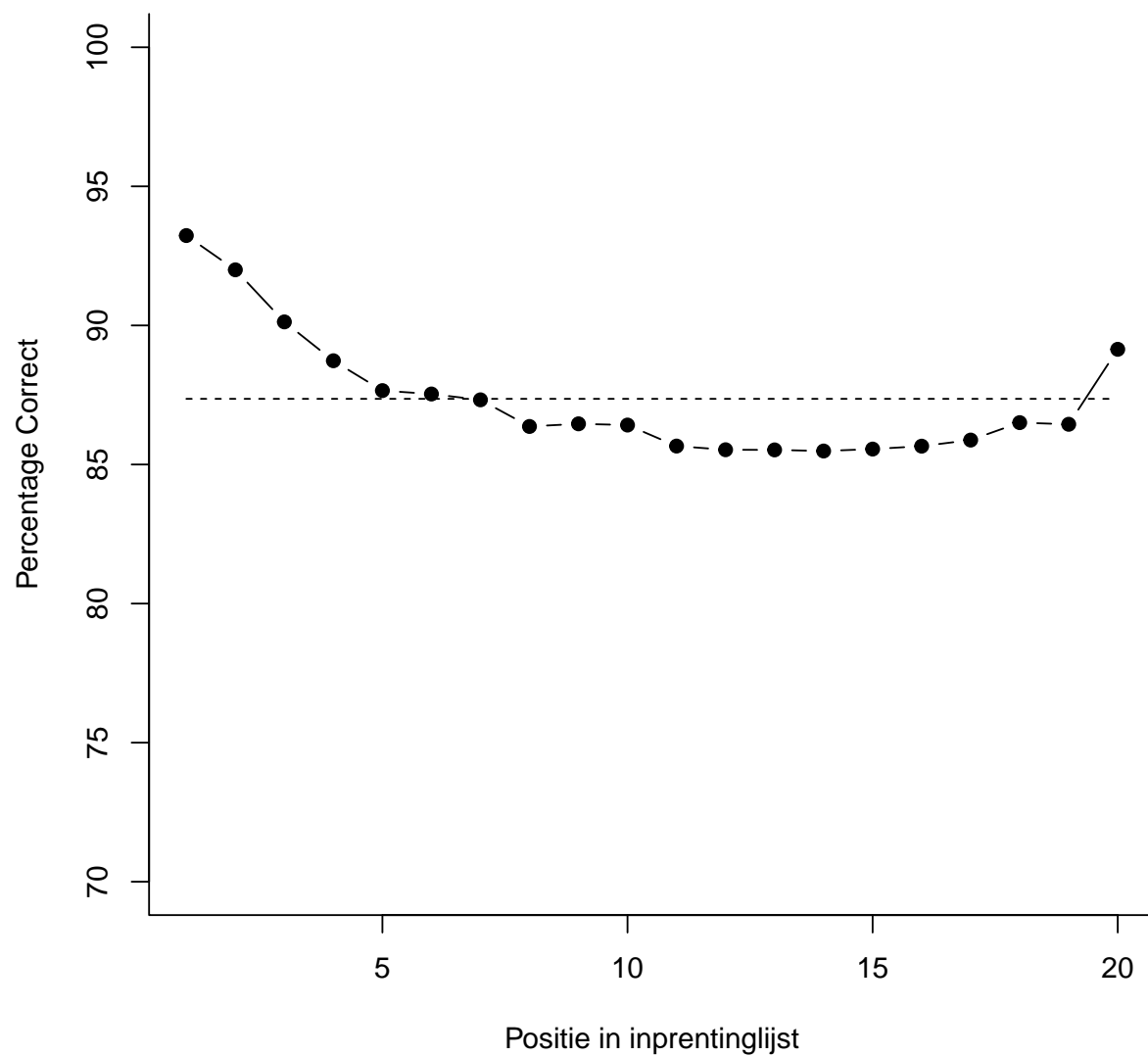


Indien we dezelfde analyse ingezoomd uitvoeren in functie van leeftijd, test, en geslacht, tekent er zich een potentieel interessant beeld af. We zien dat geslacht bij de test met beelden vooral een rol speelt tussen 20 en 60 jaar. Opnieuw is het duidelijk dat er een sterker verval is voor de woordentest dan voor de beeldentest.

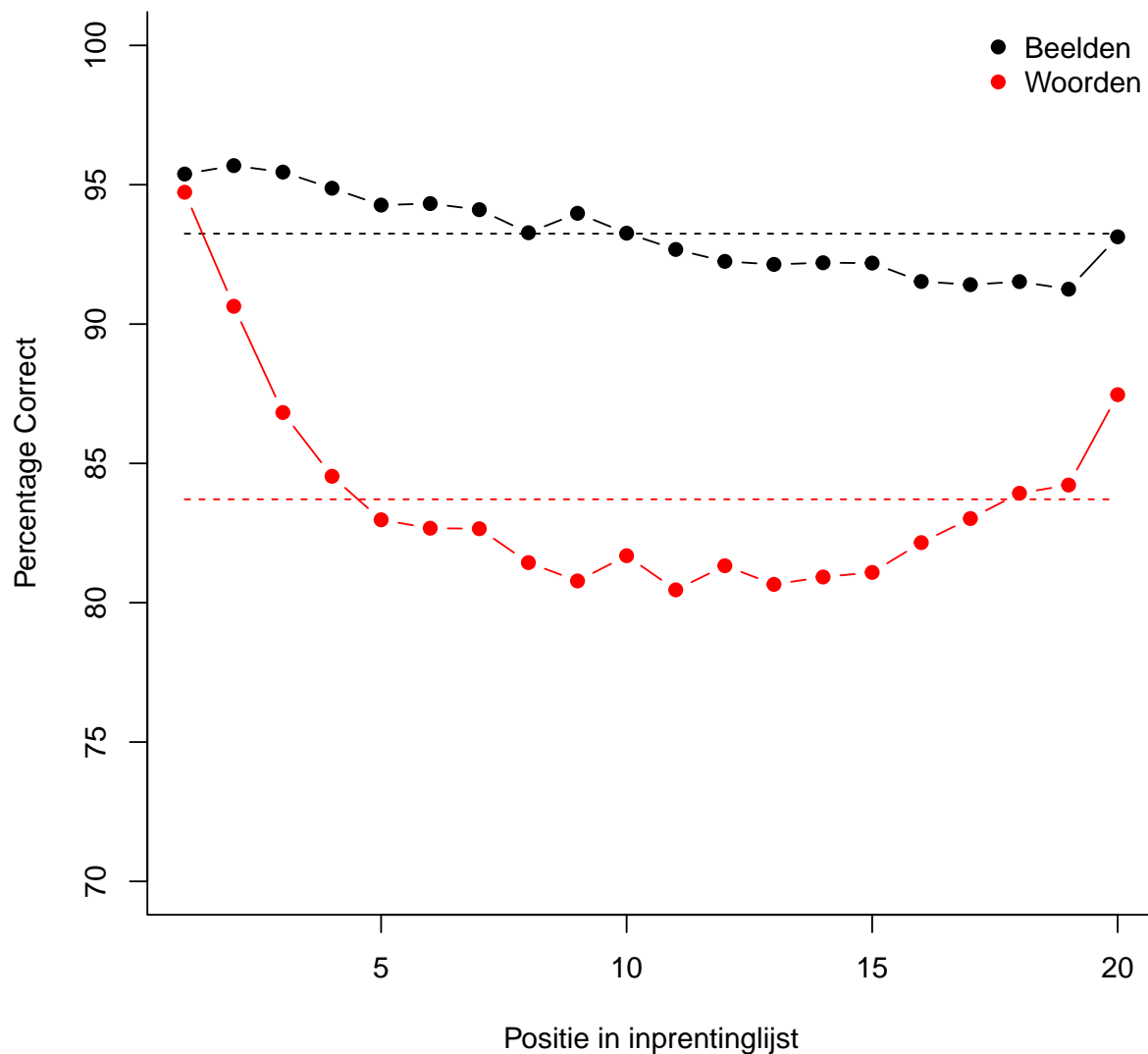


Seriële positie-effecten

Een binnen de geheugenliteratuur welbekend effect is het seriële positie-effect. Een serieel positie-effect verwijst naar een invloed van de positie waarin items aangeboden worden in de inprentingslijst op het achteraf correct herkennen van die items. Klassiek wordt er geobserveerd dat zowel eerst (primacy effect) als laatst (recency effect) aangeboden items beter herkend worden dan items in het midden van de inprentingslijst. De figuur hieronder geeft aan dat we zowel een primacy als een recency effect observeren in onze data set (over alle testen heen). Eerst aangeboden items worden beter herkend dan items in het midden, en laatst aangeboden items worden ook weer beter herkend dan items in het midden. De stippellijn geeft de gemiddelde prestatie over alle posities weer (om aan te duiden dat in het begin die prestatie erboven ligt, er vervolgens onder duikt om er dan weer boven te liggen).

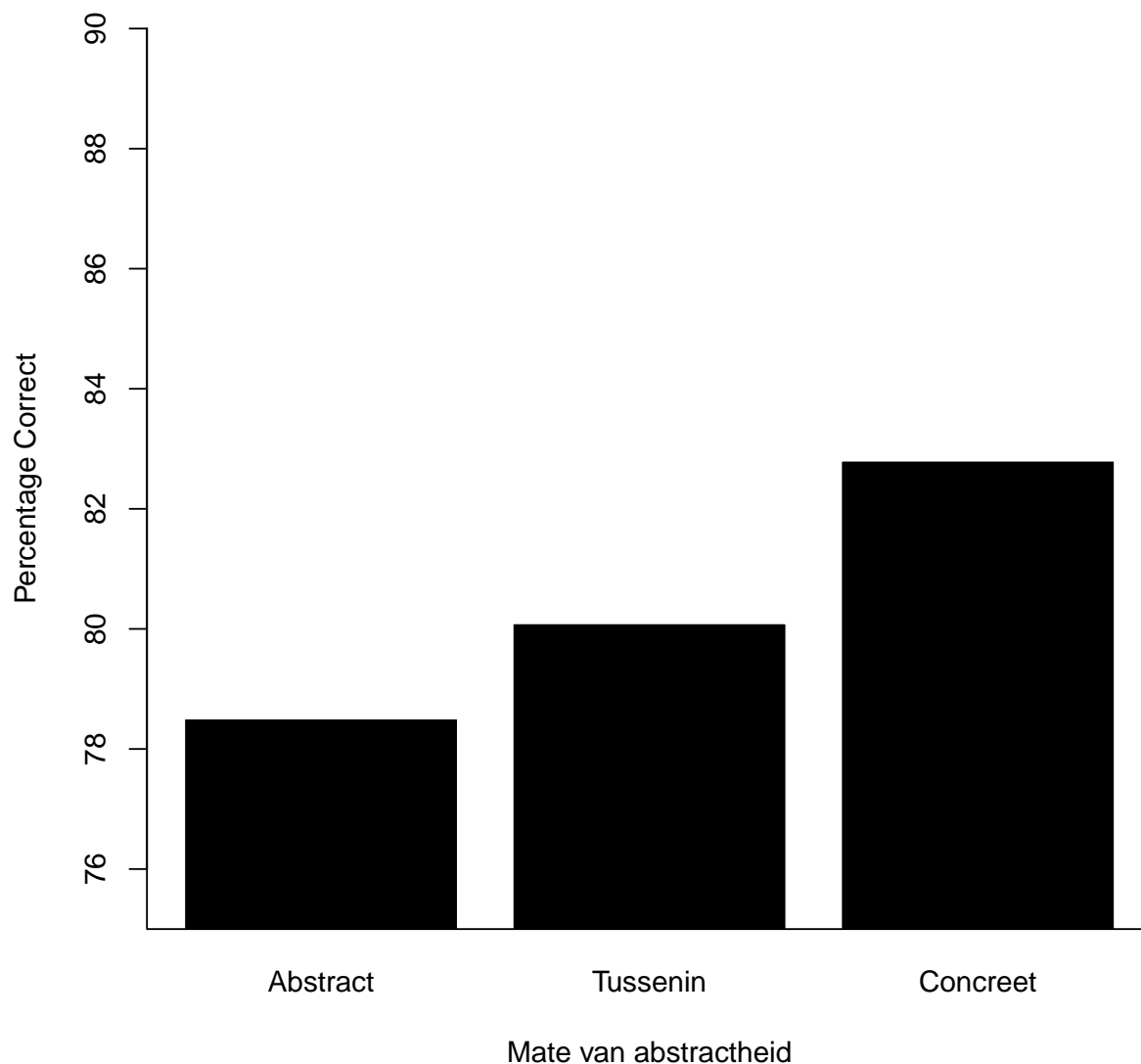


Gezien er sprake is van substantiële verschillen in performantie tussen testen, beperken we ons terug tot de twee basistesten en bekijken we eerst of er een indicatie is voor een verschil in serieel positie-effect voor de twee basistesten. Zoals blijkt uit onderstaande figuur, wordt het curvilineair patroon van bovenstaande serieel positie-effect voornamelijk veroorzaakt door de woordentest en is er eerder sprake van een strikt dalende trend voor de beeldentest.



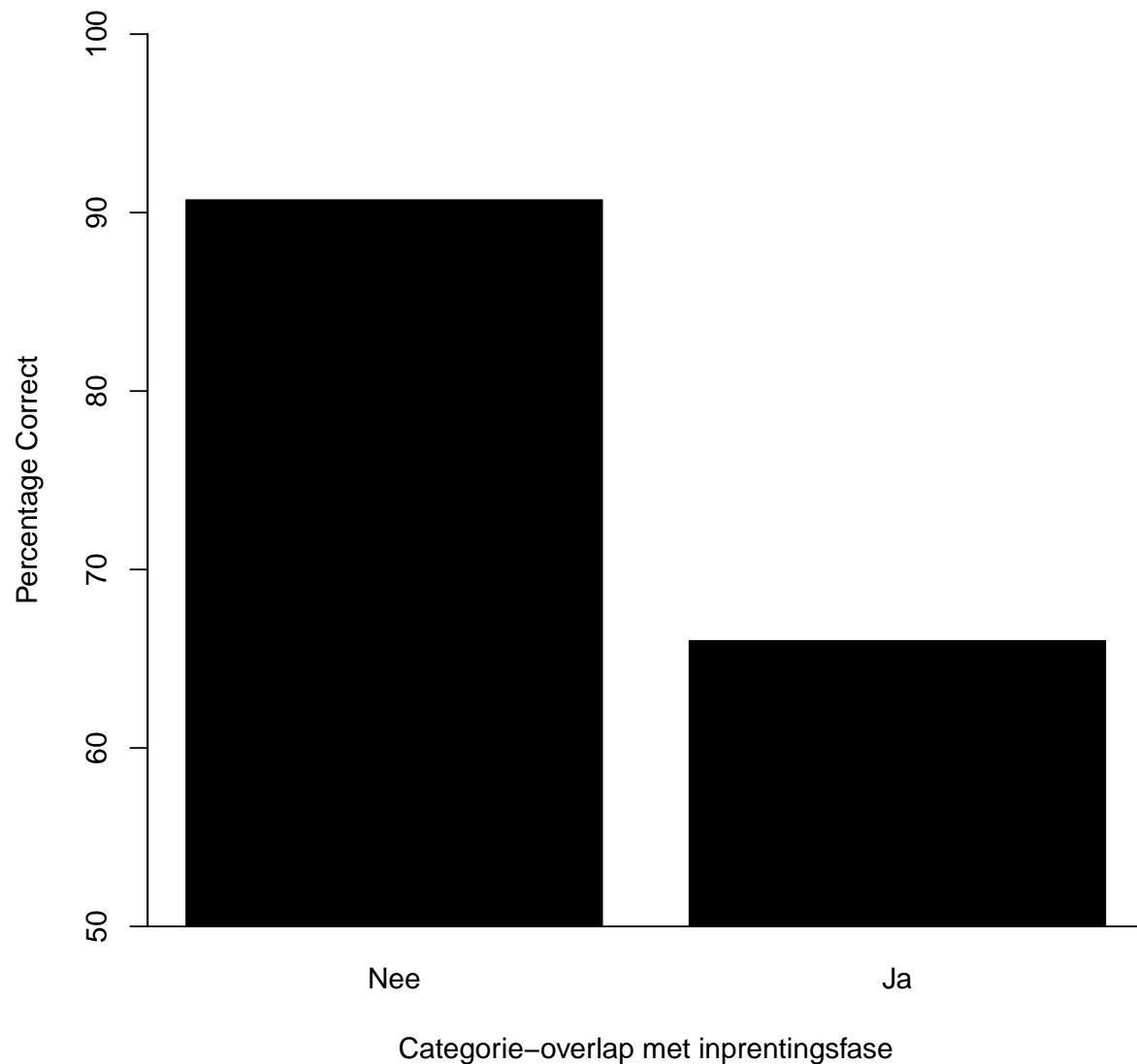
Heeft abstractheid van woorden een invloed?

In de basistest met woorden kwamen verschillende woorden voor waarvan sommige abstracter waren dan andere. De mate van abstractheid van deze woorden werd op een schaal van 1 tot 3 gecodeerd en we kunnen dus bekijken of abstractheid van een woord een invloed heeft op het percentage correct. Zoals in onderstaand staafdiagram duidelijk wordt, zijn concrete woorden inderdaad geassocieerd met een hoger percentage correct dan abstracte woorden.



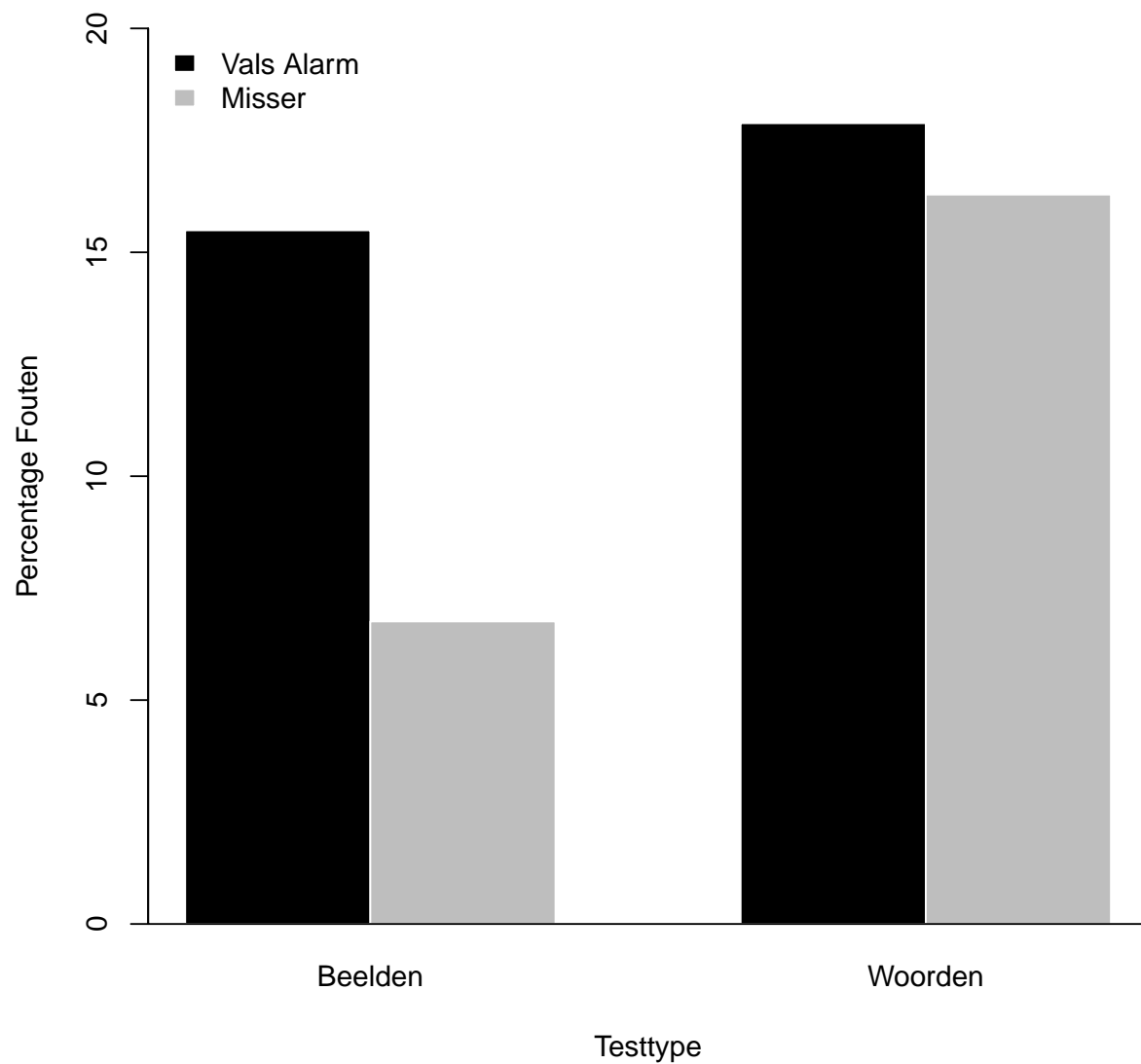
De invloed van categorie-overlap in verschillende scenes

We hebben reeds geobserveerd dat de positie van een scene in de inprentingsreeks geen sterke invloed heeft op het achteraf correct identificeren. Een aspect dat wel een invloed zou kunnen hebben op het percentage correct is de categorie van de scenes in de oproepfase vergeleken met de categorieën in de inprentingsfase. Hiermee bedoelen we dat het mogelijk is dat scenes die niet in de inprentingsfase aangeboden werden in de oproepfase minder kans hebben op correcte identificatie indien de categorie van de aangeboden scene voorkwam in de inprentingsfase. Onderstaand staafdiagram toont aan dat voor scenes die niet aangeboden werden in de inprentingsfase er een bijna maximaal percentage correct is voor de categorieën die niet aangeboden werden in de inprentingsfase. Dit percentage daalt opmerkelijk (met ongeveer 20%) voor de scenes behorend tot een categorie die wel aangeboden werd in de inprentingsfase.

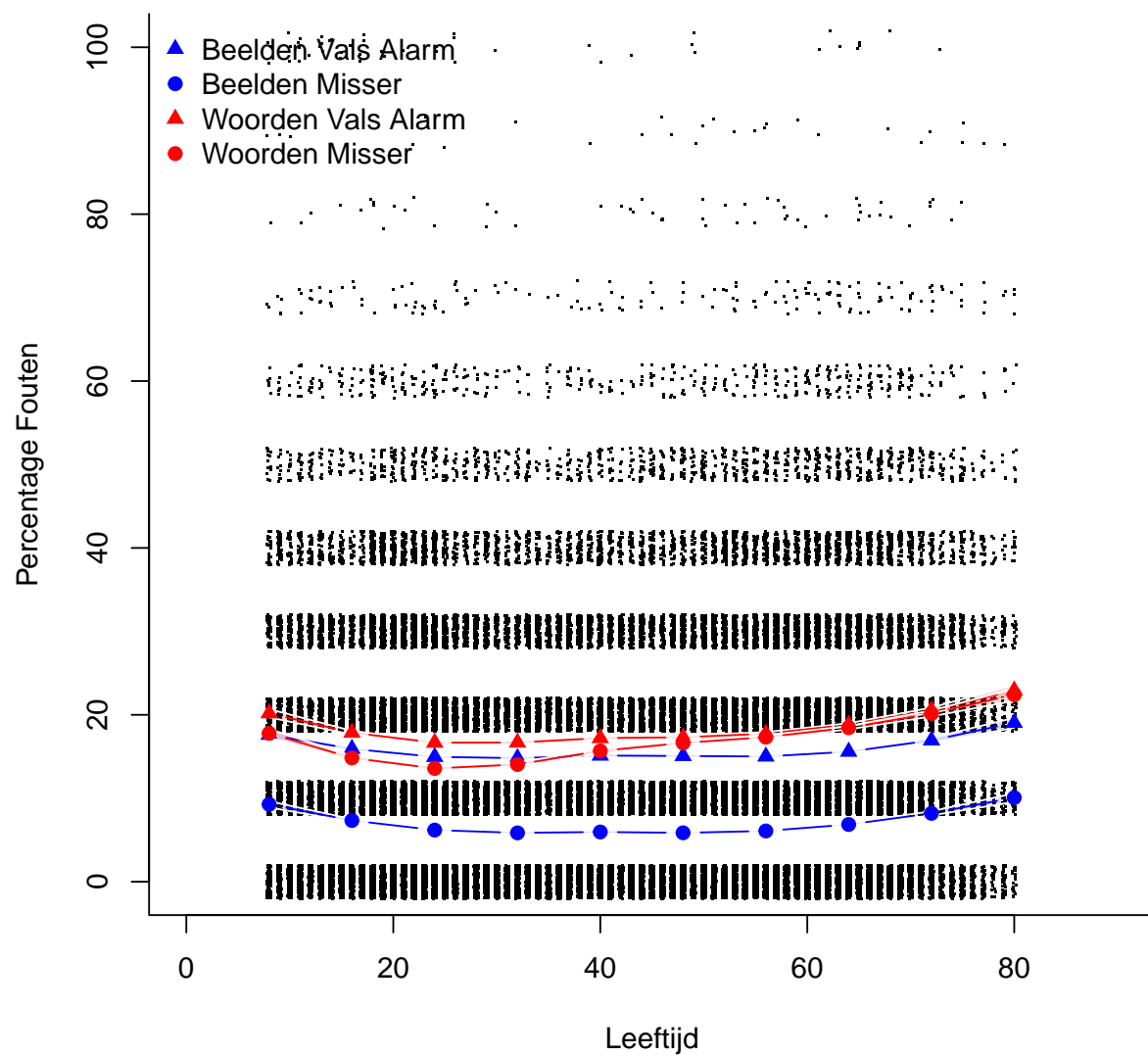


Analyse van de fouten

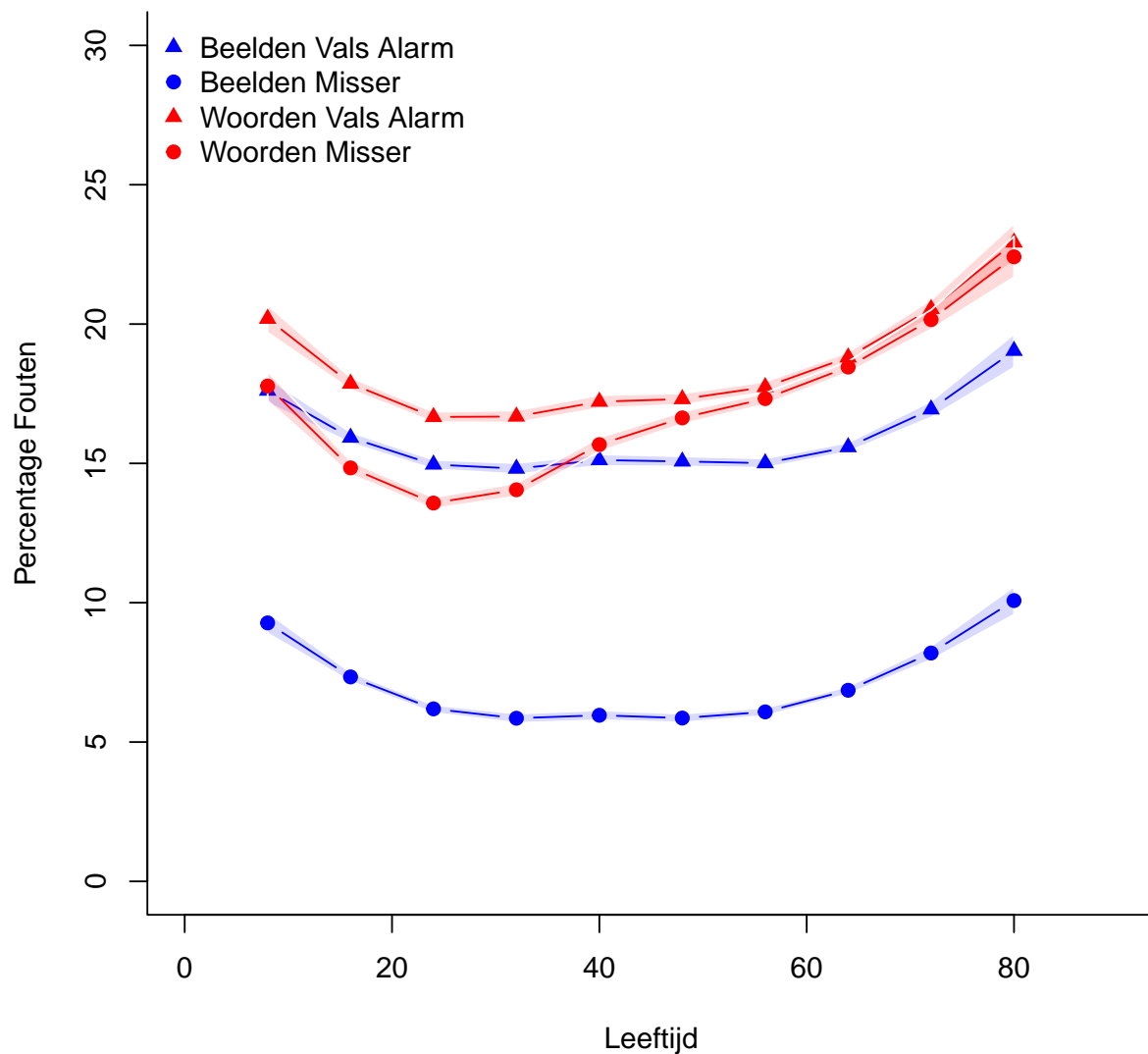
Deelnemers kunnen twee types van foute antwoorden geven. Ze kunnen enerzijds aangeven dat een item dat niet in de inprentingsfase aangeboden werd er wel in zat (vals alarm) of ze kunnen een item classificeren als niet aangeboden wanneer het wel in de inprentingsfase voorkwam(misser). Hieronder bekijken we kort of er voor de twee basistesten meer vals alarmen dan wel missers gemaakt werden en of er een verschil is tussen beide testen. Zoals reeds in de vorige analyse aangetoond werd, maken deelnemers voor beelden vooral fouten in de vals alarm categorie. Dit wordt ook duidelijk in onderstaand staafdiagram. Er is een groot verschil in percentage vals alarmen en missers voor de basistest met beelden. Voor de basistest met woorden zijn de percentages missers en vals alarmen gelijkaardig.



Net zoals we de performantie in termen van proportie correct konden analyseren in functie van leeftijd, kunnen we dat ook voor de verschillende types fouten die gemaakt worden. Op het scatterdiagram hieronder zijn alle datapunten afgebeeld voor de verschillende types fouten in functie van leeftijd. Hetzelfde gemiddeld patroon tekent zich ook af in functie van leeftijd, er worden meer fouten gemaakt voor woorden dan beelden, er is geen verschil tussen missers en vals alarmen voor woorden, waar dat wel uitgesproken is voor beelden.



In onderstaande figuur zoomen we in op het gemiddeld patroon in functie van leeftijd. Eenzelfde curvilineair patroon in functie van leeftijd tekent zich af als voor de algemene performantie.



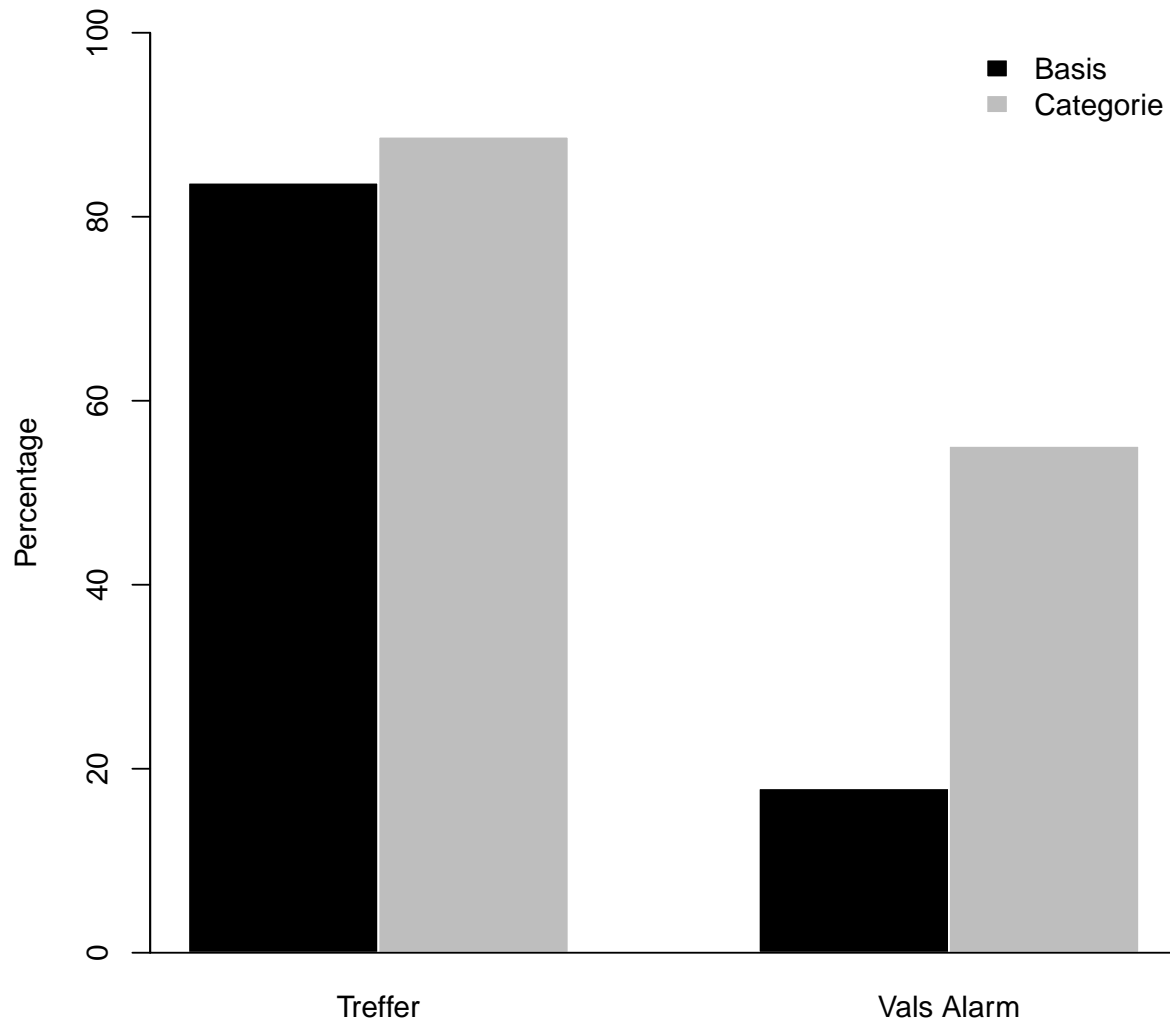
Analyse van de optionele testen

Naast de twee basistesten met beelden en woorden, waren er ook een zevental optionele testen beschikbaar. In de beschrijving van de dataset werd reeds duidelijk dat de deelnamegraad voor deze optionele testjes sterk lager lag dan voor de basistesten. We beperken ons dus tot enkele korte analyses voor de verschillende tests.

De woorden (categorie) test

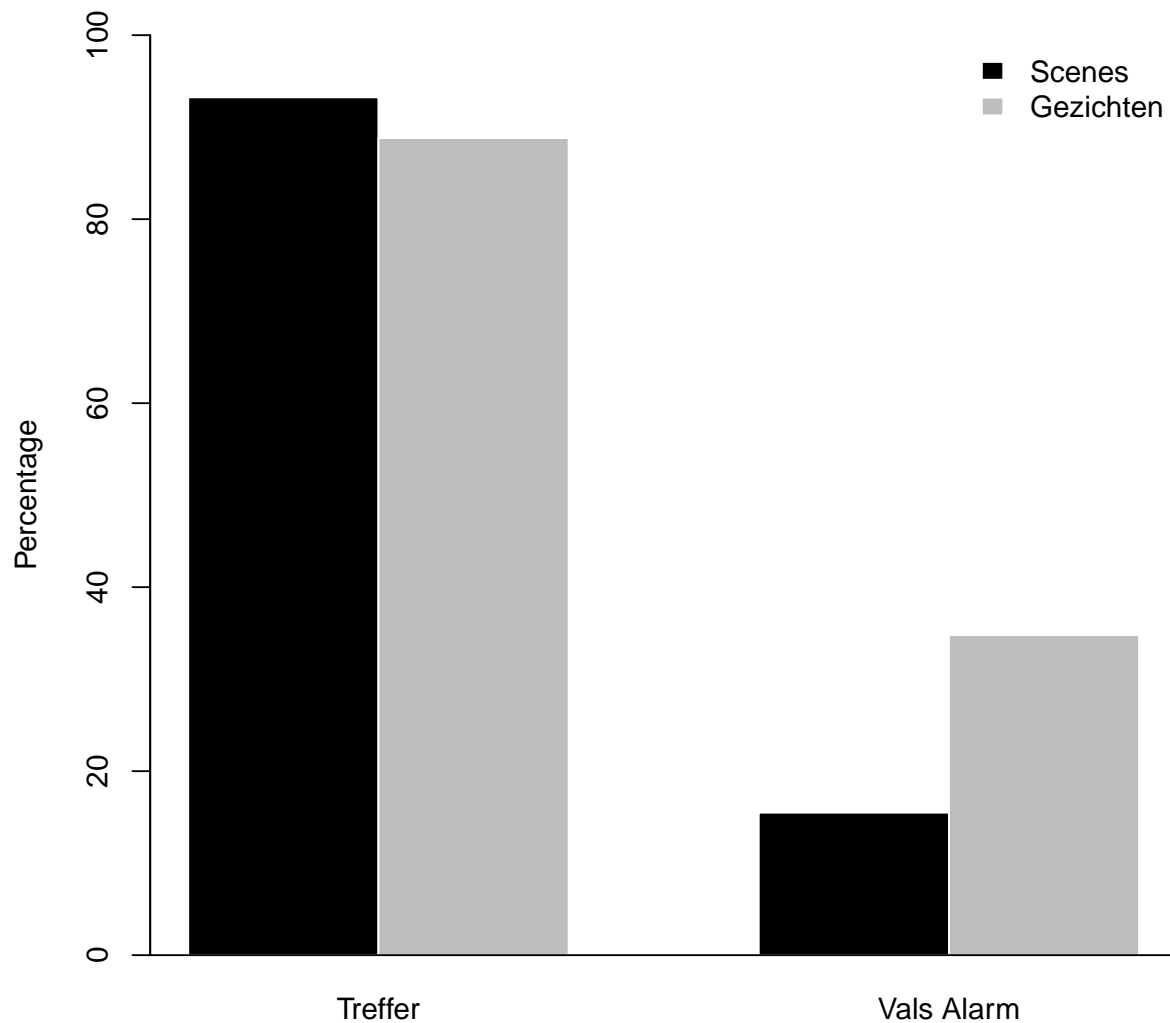
In de woorden (categorie) test werden woorden aangeboden die behoorden tot de categorie fruit dan wel groenten. De verwachting is hier dat deelnemers vooral fouten zullen maken in termen van vals alarmeren. Hiermee bedoelen we dat het meermaals presenteren van woorden uit dezelfde categorie er toe zal leiden

dat deelnemers woorden die niet in de inprentingsfase aangeboden werden zullen verwarren met aangeboden woorden en hierdoor deze woorden vaker incorrect als aangeboden zullen classificeren. In onderstaand staafdiagram zien we dat deelnemers woorden uit de inprentingsfase vaker correct classificeren dan in de basistest met woorden (treffers), maar het lager algemeen percentage correct voor de woorden (categorie) test wordt duidelijk veroorzaakt door een hoger percentage vals alarmen in deze test vergelijken met de basistest met woorden.



De beelden (gezichten) test

In de subtest met gezichten tekent zich ongeveer hetzelfde patroon af als in de categorie subtest. Het algemeen percentage correct is lager dan in de basistest met beelden hoewel het percentage treffers (gezichten aangeboden in de inprentingsfase die als dusdanig geclassificeerd werden) niet sterk verschilt van de basistest. Opnieuw ligt het percentage vals alarmen hier veel hoger dan in de basistest.



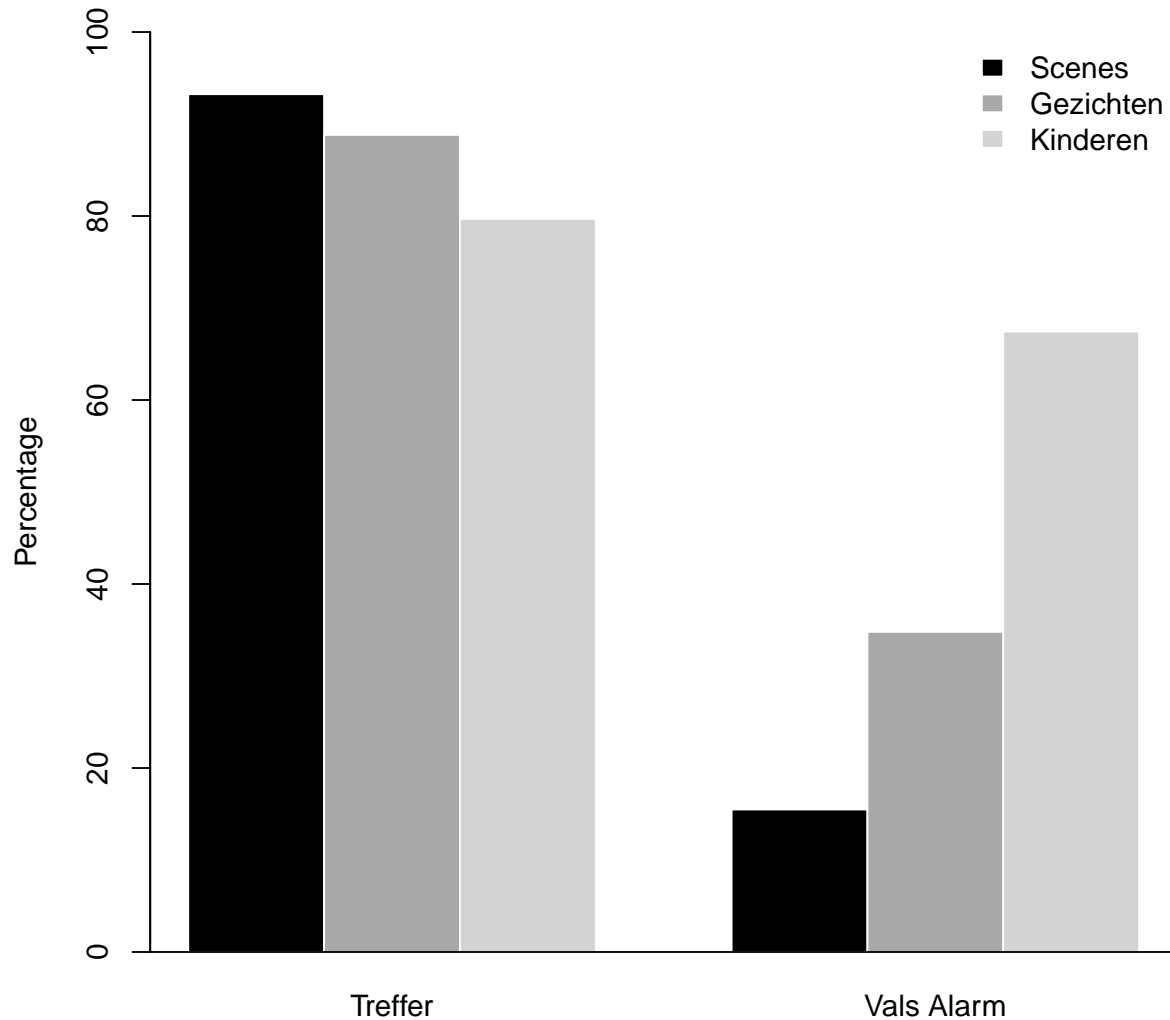
Het is mogelijk dat mannelijke deelnemers minder fouten maken voor mannelijke gezichten en vice versa voor vrouwelijke deelnemers. De tabel hieronder geeft echter weer dat dit niet het geval is. De gemiddelde percentages correct verschillen niet voor vrouwen en enkel een klein beetje voor mannen.

Deelnemer	Mannelijk gezicht	Vrouwelijk gezicht
Man	75	77
Vrouw	78	77

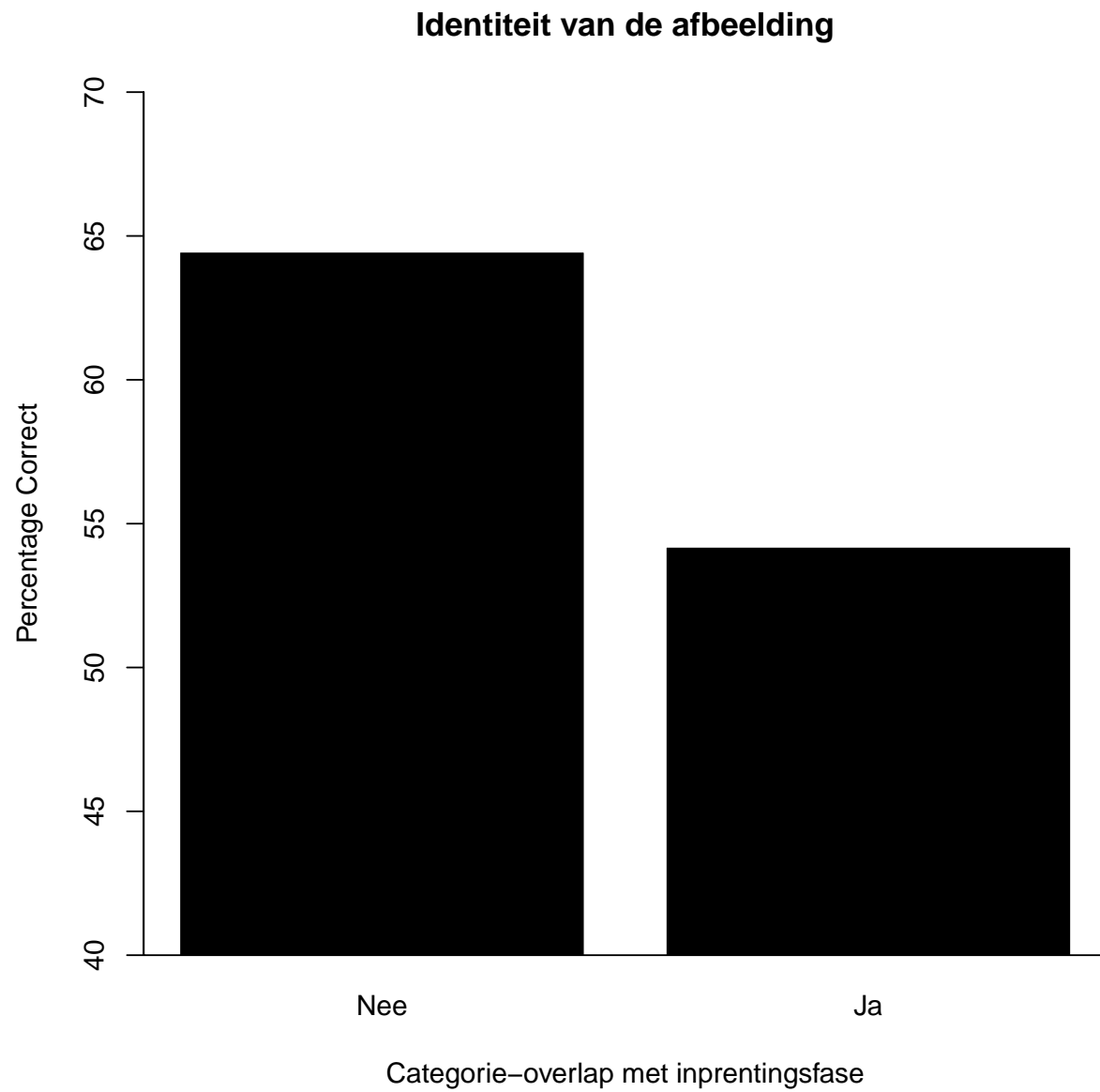
We kunnen een gelijkaardige analyse doen voor de overlap tussen de leeftijd van de deelnemers en de leeftijd van de gezichten die ze moeten herkennen. Opnieuw lijkt het effect hier beperkt en zijn deelnemers een klein beetje beter voor gezichten binnen eenzelfde leeftijdscategorie versus een verschillende leeftijdscategorie (80% vs. 81% correct).

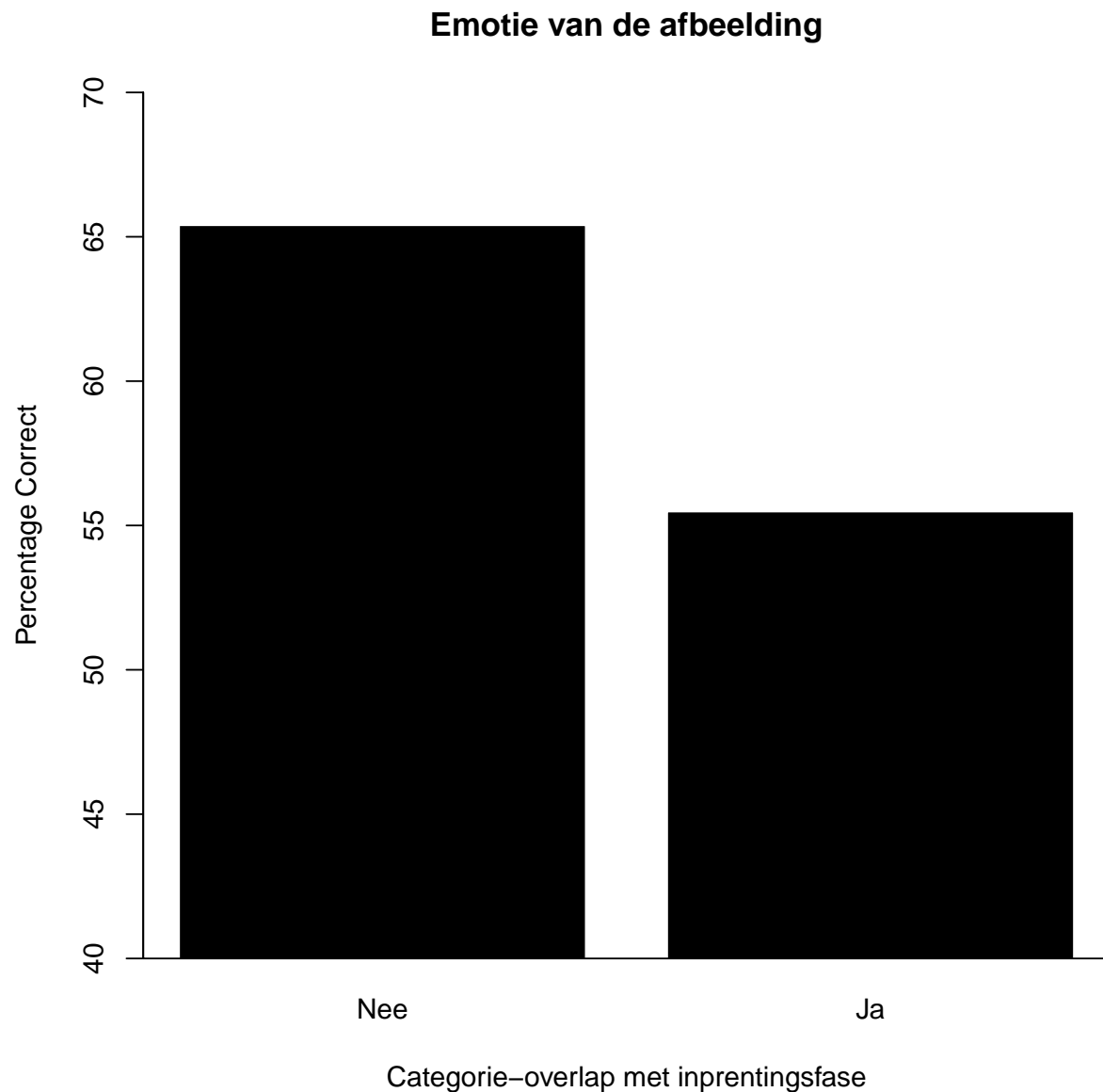
De beelden (kinderen) test

De test met afbeeldingen van kinderen is over het algemeen één van de moeilijkste testen (op die met de abstracte vormen na). Vergeleken met de basistest met beelden en de test met gezichten, daalt hier de proportie treffers ook wat, maar het lage percentage correct is vooral te wijten aan het percentage vals alarmen. Deelnemers classificeren dus zeer vaak een gezicht als reeds aangeboden hoewel het niet in de inprentingsreeks aangeboden is.



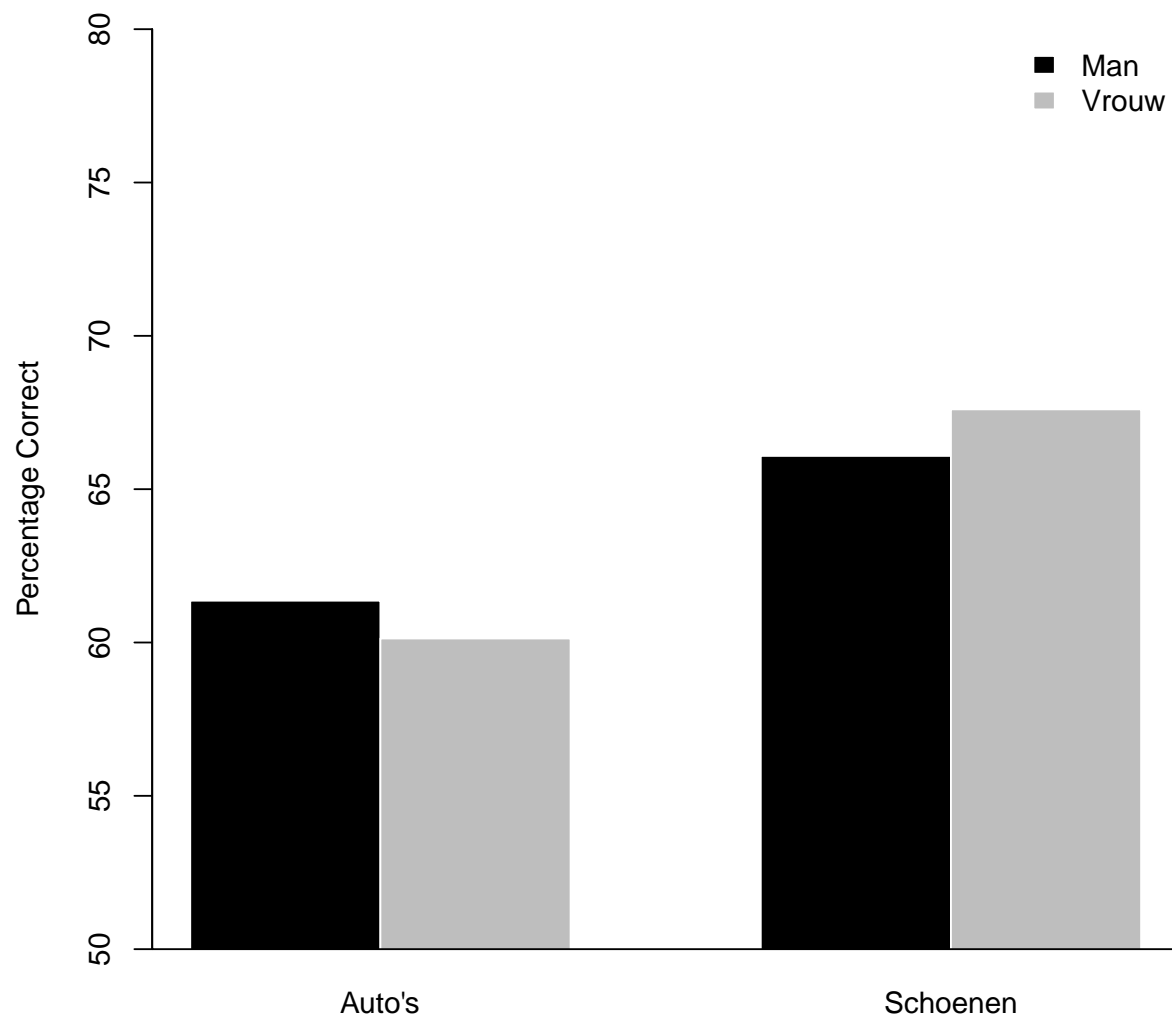
Voor de test met kinderen kunnen we net zoals bij de basistest met beelden nagaan of de performantie voor beelden die niet in de inprentingsfase aangeboden werden beïnvloed wordt door overlap categorie-overlap met beelden uit de inprentingsfase. De twee categorieën die we hier beschouwen zijn de identiteit van de afbeeldingen alsook de emoties van de afbeeldingen. Voor beide aspecten observeren we eenzelfde patroon als in de basistest met scènes. Indien er categorie-overlap is, is het percentage correct opmerkelijk lager.





De beelden (auto's) en beelden (schoenen) tests

De test met auto's en die met schoenen waren beiden moeilijker dan de basistest met scenes. De test met auto's bleek voor zowel mannen als vrouwen nog een tikkeltje moeilijker dan die met schoenen. In onderstaand staafdiagram observeren we ook een potentieel interessant effect (zij het zeer subtiel). Mannen zijn namelijk gemiddeld genomen beter op de taak met auto's daar waar vrouwen gemiddeld genomen beter zijn op de taak met schoenen.



De beelden (vormpjes) test

De moeilijkste test was die met de abstracte vormpjes. Hier scoorden deelnemers gemiddeld genomen net hoger dan 50% correct. Wanneer we de performantie op deze test vergelijken met die op de basistest met beelden wordt duidelijk dat de lage performantie niet enkel te wijten is aan een hoog percentage vals alarmen, maar ook aan een lager percentage treffers. In die zin wijkt de performantie op deze test dus af van de performantie op de andere testen met beelden. Onderstaand staafdiagram vergelijkt de vormpjes test met de scenes test in termen van percentages treffers en vals alarmen.

